RECOMBINANT PHYTASES AND USES THEREOF

FIELD OF THE INVENTION

This invention relates to newly made polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production and isolation of such polynucleotides and polypeptides. More particularly, the polypeptides of the present invention have been identified as phytases and in particular, enzymes having phytase activity.

BACKGROUND

Minerals are essential elements for the growth of all organisms. Dietary minerals can be derived from many source materials, including plants. *E.g.*, plant seeds are a rich source of minerals since they contain ions that are complexed with the phosphate groups of phytic acid molecules. These phytate-associated minerals satisfy the dietary needs of some species of farmed organisms, such as multi-stomached ruminants. Accordingly, ruminants do not require dietary supplementation with inorganic phosphate and minerals because microorganisms in the rumen produce enzymes that catalyze conversion of phytate (myo-inositol-hexaphosphate) to inositol and inorganic phosphate. In the process, minerals that have been complexed with phytate are released. The majority of species of farmed organisms, however, are unable to efficiently utilize phytate-associated minerals. Thus, for example, in the livestock production of monogastric animals (*e.g.*, pigs, birds, and fish), feed is commonly supplemented with minerals and/or with antibiotic substances that alter the digestive flora environment of the consuming organism to enhance growth rates.

As such, there are many problematic burdens - related to nutrition, *ex vivo* processing steps, health and medicine, environmental conservation, and resource management – that are associated with an insufficient hydrolysis of phytate in many applications. The following are non-limiting examples of these problems:

1) The supplementation of diets with inorganic minerals is a costly expense.
2) The presence of unhydrolyzed phytate is undesirable and problematic in many ex vivo applications (*e.g.* by causing the presence of unwanted sludge).

3) The supplementation of diets with antibiotics poses a medical threat to humans and animals alike by increasing the abundance of antibiotic-tolerant pathogens.

4) The discharge of unabsorbed fecal minerals into the environment disrupts and damages the ecosystems of surrounding soils, fish farm waters, and surface waters at large.

5) The valuable nutritional offerings of many potential foodstuffs remain significantly untapped and squandered.

Many potentially nutritious plants, including particularly their seeds, contain appreciable amounts of nutrients, *e.g.* phosphate, that are associated with phytate in a manner such that these nutrients are not freely available upon consumption. The unavailability of these nutrients is overcome by some organisms, including cows and other ruminants, that have a sufficient digestive ability - largely derived from the presence of symbiotic life forms in their digestive tracts – to hydrolyze phytate and liberate the associated nutrients. However, the majority of species of farmed animals, including pigs, fish, chickens, turkeys, as well as other non-ruminant organisms including man, are unable to efficiently liberate these nutrients after ingestion.

Consequently, phytate-containing foodstuffs require supplementation with exogenous nutrients and/or with a source of phytase activity in order to ammend their deficient nutritional offerings upon consumption by a very large number of species of organisms.

In yet another aspect, the presence of unhydrolized phytate leads to problematic consequences in *ex vivo* processes including - but not limited to - the processing of foodstuffs. In but merely one exemplification, as described in EP0321004-B1 (Vaara *et al.*), there is a step in the processing of corn and sorghum kernels whereby the hard kernels are steeped in water to soften them. Water-soluble subtances that leach out during this process become part of a corn steep liquor, which is concentrated by evaporation. Unhydrolized phytic acid in the corn steep liquor, largely in the form of calcium and magnesium salts, is associated with phosphorus and deposits an undesirable sludge with proteins and metal ions. This sludge is problematic in the

evaporation, transportation and storage of the corn steep liquor. Accordingly, the instantly disclosed phytase molecules - either alone or in combination with other reagents (including but not limited to enzymes, including proteases) - can be used not only in this application (*e.g.*, for prevention of the unwanted slugde) but also in other applications where phytate hydrolysis is desirable.

The supplementation of diets with antibiotic substances has many beneficial results in livestock production. For example, in addition to its role as a prophylactic means to ward off disease, the administration of exogenous antibiotics has been shown to increase growth rates by upwards of 3-5%. The mechanism of this action may also involve – in part – an alteration in the digestive flora environment of farmed animals, resulting in a microfloral balance that is more optimal for nutrient absorption.

However, a significant negative effect associated with the overuse of antibiotics is the danger of creating a repository of pathogenic antibiotic-resistant microbial strains. This danger is imminent, and the rise of drug-resistant pathogens in humans has already been linked to the use of antibiotics in livestock. For example, Avoparcin, the antibiotic used in animal feeds, was banned in many places in 1997, and animals are now being given another antibiotic, virginiamycin, which is very similar to the new drug, Synercid, used to replace vancomycin in human beings. However, studies have already shown that some enterococci in farm animals are resistant to Synercid. Consequently, undesired tolerance consequences, such as those already seen with Avoparcin and vancomycin, are likely to reoccur no matter what new antibiotics are used as blanket prophylactics for farmed animals. Accordingly, researchers are calling for tighter controls on drug use in the industry.

The increases in growth rates achieved in animals raised on foodstuffs supplemented with the instantly disclosed phytase molecules matches – if not exceeds – those achieved using antibiotics such as, for example, Avoparcin. Accordingly, the instantly disclosed phytase molecules - either alone or in combination with other reagents (including but not limited to enzymes, including proteases) - are serviceable not only in this application (*e.g.*, for increasing the growth rate of farmed animals) but also in other applications where phytate hydrolysis is desirable.

An environmental consequence is that the consumption of phytate-containing foodstuffs by any organism species that is phytase-deficient - regardless of whether the foodstuffs are supplemented with minerals - leads to fecal pollution resulting from the excretion of unabsorbed minerals. This pollution has a negative impact not only on the immediate habitat but consequently also on the surrounding waters. The environmental alterations occur primarily at the bottom of the food chain, and therefore have the potential to permeate upwards and throughout an ecosystem to effect permanent and catastrophic damage – particularly after years of continual pollution. This problem has the potential to manifest itself in any area where concentrated phytate processing occurs – including *in vivo* (*e.g.* by animals in areas of livestock production, zoological grounds, wildlife refuges, etc.) and *in vitro* (*e.g.* in commercial corn wet milling, ceral steeping processes, and the like) processing steps.

The decision to use exogenously added phytase molecules – whether to fully replace or to augment the use of exogenously administered minerals and/or antibiotics – ultimately needs to pass a test of financial feasibility and cost effectiveness by the user whose livelihood depends on the relevant application, such as livestock production.

Consequently, there is a need for means to achieve efficient and cost effective hydrolysis of phytate in various applications. Particularly, there is a need for means to optimize the hyrolysis of phytate in commercial applications. In a particular aspect, there is a need to optimize commercial treatment methods that improve the nutritional offerings of phytate-containing foodstuffs for consumption by humans and farmed animals.

Previous reports of recombinant phytases are available, but their inferior activities are eclipsed by the newly discovered phytase molecules of instant invention. Accordingly, the instantly disclosed phytase molecules provide substantially superior commercial performance than previously identified phytase molecules, *e.g.* phytase molecules of fungal origin.

Phytate occurs as a source of stored phosphorous in virtually all plant feeds (Graf (Ed.), 1986). Phytic acid forms a normal part of the seed in cereals and legumes. It functions to bind dietary minerals that are essential to the new plant as it emerges from the seed. When the phosphate groups of phytic acid are removed by the seed enzyme phytase, the ability to bind metal ions is lost and the minerals become available to the plant. In livestock feed grains, the trace minerals bound by phytic acid are largely unavailable for absorption by monogastric animals, which lack phytase activity.

Although some hydrolysis of phytate occurs in the colon, most phytate passes through the gastrointestinal tract of monogastric animals and is excreted in the manure contributing to fecal phosphate pollution problems in areas of intense livestock production. Inorganic phosphorous released in the colon has an appreciably diminished nutritional value to livestock because inorganic phosphorous is absorbed mostly – if not virtually exclusively - in the small intestine. Thus, an appreciable amount of the nutritionally important dietary minerals in phytate is unavailable to monogastric animals.

In sum, phytate-associated nutrients are comprised of not only phosphate that is covalently linked to phytate, but also other minerals that are chelated by phytate as well. Moreover, upon injestion, unhydrolyzed phytate may further encounter and become associated with additional minerals. The chelation of minerals may inhibit the activity of enzymes for which these minerals serve as co-factors.

Conversion of phytate to inositol and inorganic phosphorous can be catalyzed by microbial enzymes referred to broadly as phytases. Phytases such as phytase #EC 3.1.3.8 are capable of catalyzing the hydrolysis of myo-inositol hexaphosphate to D-myo-inositol 1,2,4,5,6-pentaphosphate and orthophosphate. Certain fungal phytases reportedly hydrolyze inositol pentaphosphate to tetra-, tri-, and lower phosphates. For example, *A. ficuum* phytases reportedly produce mixtures of myoinositol di- and mono-phosphates (Ullah, 1988). Phytase-producing microorganisms are comprised of bacteria such as *Bacillus subtilis* (Powar and Jagannathan, 1982) and Pseudomonas (Cosgrove, 1970); yeasts such as *Sacchoromyces cerevisiae* (Nayini and Markakis, 1984); and fungi such as *Aspergillus terreus* (Yamada *et al.*, 1968).

Acid phosphatases are enzymes that catalytically hydrolyze a wide variety of phosphate esters and usually exhibit pH optima below 6.0 (Igarashi and Hollander, 1968). For example, #EC 3.1.3.2 enzymes catalyze the hydrolysis of orthophosphoric monoesters to orthophosphate products. An acid phosphatase has reportedly been purified from *A. ficuum*. The deglycosylated form of the acid phosphatase has an apparent molecular weight of 32.6 kDa (Ullah *et al.*, 1987).

Phytase and less specific acid phosphatases are produced by the fungus *Aspergillus ficuum* as extracellular enzymes (Shieh *et al.*, 1969). Ullah reportedly purified a phytase from wild-type *A. ficuum* that had an apparent molecular weight of 61.7 kDA (on SDS-PAGE; as corrected for glycosylation); pH optima at pH 2.5 and pH 5.5; a Km of about 40μm; and, a specific activity of about 50 U/mg (Ullah, 1988). PCT patent application WO 91/05053 also reportedly discloses isolation and molecular cloning of a phytase from *Aspergillus ficuum* with pH optima at pH 2.5 and pH 5.5, a Km of about 250 μm, and specific activity of about 100 U/mg protein.

Summarily, the specific activity cited for these previously reported microbial enzymes has been approximately in the range of 50-100 U/mg protein. In contrast, the phytase activity disclosed in the instant invention has been measured to be approximately 4400 U/mg. This corresponds to about a 40-fold or better improvement in activity.

The possibility of using microbes capable of producing phytase as a feed additive for monogastric animals has been reported previously (U.S. Patent No. 3,297,548 Shieh and Ware; Nelson *et al.*, 1971). The cost-effectiveness of this approach has been a major limitation for this and other commercial applications. Therefore improved phytase molecules are highly desirable.

Microbial phytases may also reportedly be useful for producing animal feed from certain industrial processes, *e.g.*, wheat and corn waste products. In one aspect, the wet milling process of corn produces glutens sold as animal feeds. The addition of phytase may reportedly improve the nutritional value of the feed product. For example,

the use of fungal phytase enzymes and process conditions (t~50°C and pH ~5.5) have been reported previously in (*e.g.* EP 0 321 004). Briefly, in processing soybean meal using traditional steeping methods, *i.e.*, methods without the addition of exogenous phytase enzyme, the presence of unhydrolyzed phytate reportedly renders the meal and wastes unsuitable for feeds used in rearing fish, poultry and other non-ruminants as well as calves fed on milk. Phytase is reportedly useful for improving the nutrient and commercial value of this high protein soy material (see Finase Enzymes by Alko, Rajamäki, Finland). A combination of fungal phytase and a pH 2.5 optimum acid phosphatase form *A. niger* has been used by Alko, Ltd as an animal feed supplement in their phytic acid degradative product Finas F and Finase S. However, the cost-effectiveness of this approach has remained a major limitation to more widespread use. Thus a cost-effective source of phytase would greatly enhance the value of soybean meals as an animal feed (Shieh *et al.*, 1969).

To solve the problems disclosed, the treatment of foodstuffs with exogenous phytase enzymes has been proposed, but this approach was not been fully optimized, particularly with respect to feasibility and cost efficiency. This optimization requires the consideration that a wide range of applications exists, particularly for large scale production. For example, there is a wide range of foodstuffs, preparation methods thereof, and species of recipient organisms.

In a particular exemplification, it is appreciated that the manufacture of fish feed pellets requires exposure of ingedients to high temperatures and/or pressure in order to produce pellets that do not dissolve and/or degrade prematurely (*e.g.* prior to consumption) upon subjection to water. It would thus be desirable for this manufacturing process to obtain additive enzymes that are stable under high temperature and/or pressure conditions. Accordingly it is appreciated that distinct phytases may be differentially preferable or optimal for distinct applications.

It is furthermore recognized that an important way to optimize an enzymatic process is through the modification and improvement of the pivotal catalytic enzyme. For example, a transgenic plant can be formed that is comprised of an expression system for expressing a phytase molecule. It is appreciated that by attempting to

improve factors that are not directly related to the activity of the expressed molecule proper, such as the expression level, only a finite - and potentially insufficient - level of optimization may be maximally achieved. Accordingly, there is also a need for obtaining molecules with improved characteristics.

A particular way to achieve improvements in the characteristics of a molecule is through a technological approach termed directed evolution, including Diversa Corporation's proprietary approaches for which the term DirectEvolution® has been coined and registered. These approaches are further elaborated in Diversa's co-owned patent (US 5,830,696) as well as in several co-pending patent applications. In brief, DirectEvolution® comprises: a) the subjection of one or more molecular template to mutagenesis to generate novel molecules, and b) the selection among these progeny species of novel molecules with more desirable characteristics.

However, the power of directed evolution depends on the starting choice of starting templates, as well as on the mutagenesis process(es) chosen and the screening process(es) used. For example, the approach of generating and evaluating a full range of mutagenic permutations on randomly chosen molecular templates and/or on initial molecular templates having overly suboptimal properties is often a forbiddingly large task. The use of such templates offers, at best, a circuitously suboptimal path and potentially provides very poor prospects of yielding sufficiently improved progeny molecules. Additionally, it is appreciated that our current body of knowledge is very limited with respect to the ability to rigorously predict beneficial modifications.

Consequently, it is a desirable approach to discover and to make use of molecules that have pre-evolved properties – preferably pre-evolved enzymatic advantages – in nature. It is thus appreciated in the instant disclosure that nature provides (through what has sometimes been termed "natural evolution") molecules that can be used immediately in commercial applications, or that alternatively, can be subjected to modifications, such as directed, evolution to achieve even greater improvements.

In sum, there is a need for novel, highly active, physiologically effective, and economical sources of phytase activity. Specifically, there is a need to identify novel phytases that: a) have superior activities under one or more specific applications, and are thus useful for optimizing these specific applications; b) are useful as templates for directed evolution to achieve even further improved novel molecules; and c) are useful as tools for the identification of additional related molecules by means such as hybridization-based approaches. This invention meets these needs in a novel way.

SUMMARY OF THE INVENTION

In a first aspect, the invention provides an isolated nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:1, the complement of SEQ ID NO:1, SEQ ID NO:3, the complement of SEQ ID NO:3, SEQ ID NO:5, the complement of SEQ ID NO:5, SEQ ID NO:7, the complement of SEQ ID NO:7, SEQ ID NO:9, the complement of SEQ ID NO:9, SEQ ID NO:11, the complement of SEQ ID NO:11, SEQ ID NO:13, and the complement of SEQ ID NO:13.

In various embodiments thereof, the nucleic acid is at least 95% identical or at least 90% identical or at least 80% identical or at least 70% identical to a sequence of a nucleic acid of the first aspect as determined by analysis with a sequence comparison algorithm.

In other embodiments, the invention provides a nucleic acid that hybridizes to a nucleic acid of the first aspect under conditions of high stringency or under conditions of moderate stringency or under conditions of low stringency.

Embodiments of various aspects of the invention are drawn to expression vectors having the nucleic acid of the first aspect and an expression control nucleotide sequence. In other embodiments of the aspects of the invention, the invention provides a host cell transformed with the nucleic acid of the invention or a host cell transformed with the an expression vector of the invention.

In a second asepct, the invention provides a nucleotide sequence encoding a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14.

In a third aspect, the invention provides an isolated nucleic acid comprising a nucleotide sequence encoding a polypeptide having at least thirty contiguous amino acids of a protein having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14.

In a fourth aspect, the invention provides an isolated phytase protein comprising a polypeptide having at least thirty contiguous amino acids of a protein having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14.

In a fifth aspect, the invention provides an isolated phytase protein comprising a polypeptide having at least thirty contiguous amino acids of a protein having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14, wherein the SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14 have at least one conservative amino acid substitution.

In sixth aspect, the invention provides a nucleic acid expression vector. The expression vector comprises a nucleotide sequence encoding a polypeptide having at least thirty contiguous amino acids of a protein having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14; and an expression control nucleotide sequence.

In various embodiments of this aspect, the invention provides a nucleic acid expression vector in which the expression control nucleotide sequence is a constitutive promoter or the expression control nucleotide sequence is a tissue-specific promoter. In

yet other embodiments thereof, the nucleic acid expression vector includes a nucleotide sequence encoding a signal peptide. In a specific embodiment thereof, the signal peptide is the PR protein PR-S signal peptide from tobacco.

In a seventh aspect, the invention provides a method of improving the nutritional value of a phytate-containing foodstuff, the method comprising contacting the phytate-containing foodstuff with a substantially pure phytase enzyme having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14, the phytase enzyme catalyzing the liberation of inorganic phosphate from the phytate-containing foodstuff, thereby improving the nutritive value of the contacted foodstuff.

In certain embodiments of the seventh aspect, the phytase enzyme is produced by a recombinant expression system and the expression of the phytase-encoding nucleic acid results in the production of the phytase enzyme.

In certain embodiments of the seventh aspect, the invention provides method in which the liberation of the inorganic phosphate from the phytate in the phytate-containing foodstuff occurs prior to the ingestion of the phytate-containing foodstuff by a recipient organism. Alternatively, the liberation of the inorganic phosphate from the phytate in the phytate-containing foodstuff occurs after the ingestion of the phytate-containing foodstuff by a recipient organism. Alternatively, the liberation of the inorganic phosphate from the phytate in the phytate-containing foodstuff occurs in part prior to, and in part after, the ingestion of the phytate-containing foodstuff by a recipient organism.

In an eighth aspect, the invention provides a method to produce an animal feed. The method comprises transforming a plant, plant part, or plant cell with a nucleic acid expression vector of the invention, culturing the plant, plant part or plant cell under conditions in which the phytase protein is expressed, and converting the plant, plant parts, or plant cell into a composition suitable for animal feed. In some embodiments

of this aspect, the feed is designed for a monogastric animal or the feed is designed for a ruminant.

In a ninth aspect, the invention provides a non-human transgenic organism having a heterologous nucleic acid encoding a polypeptide having at least thirty contiguous amino acids of a protein having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14. In certain embodiments thereof, the non-human transgenic organism. In embodiments thereof, the heterologous nucleic acid is expressed in a seed.

In a tenth aspect, the invention provides a method of producing a substantially purified phytase protein. The method comprises expressing in a cell a phytase a polypeptide having at least thirty contiguous amino acids of a protein having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14, and recovering the phytase protein. In certain embodiments of the tenth aspect, the cell is a prokaryotic or eukaryotic cell. In other certain embodiments, the phytase protein is glycosylated.

In an eleventh aspect, the invention provides a method of increasing resistance of a phytase polypeptide to enzymatic inactivation in a digestive system of an animal, the method comprising glycosylating the phytase polypeptide. In embodiments thereof, the phytase glycosylation is N-linked glycosylation. In some embodiments thereof, the phytase polypeptide is glycosylated as a result of *in vivo* expression in a eukaryotic cell selected from the group consisting of a fungal, a plant cell, or a mammalian cell.

In a twelfth aspect, the invention provides a feed composition. The composition comprises a plant, plant part, or plant cell expressing a polypeptide having at least thirty contiguous amino acids of a protein having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14; and a phytate-containing foodstuff. In one certain embodiment thereof, the plant part is a seed or portion thereof.

In a thirteenth aspect, the invention provides a feed composition that comprises a substantially purified phytase protein having at least thirty contiguous amino acids of a protein having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14, and a phytate-containing foodstuff. In certain embodiments thereof, the feed is manufactured in pellet form and/or produced using polymer coated additives. In other certain embodiments thereof, the substantially purified phytase protein of the feed is provided in granulate form. In another embodiment of this aspect, the feed is produced by spray drying.

In a fourteenth aspect, the invention provides an antibody or fragment thereof that specifically recognizes an epitope contained in an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14. In various embodiments thereof, the antibody or fragment thereof is a polyclonal antibody or the antibody or fragment thereof is a monoclonal antibody.

In fifteenth aspect, the invention provides a method of generating a variant phytase. The method comprises obtaining a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NO:1, the complement of SEQ ID NO:1, SEQ ID NO:3, the complement of SEQ ID NO:3, SEQ ID NO:5, the complement of SEQ ID NO:5, SEQ ID NO:7, the complement of SEQ ID NO:7, SEQ ID NO:9, the complement of SEQ ID NO:9, SEQ ID NO:11, the complement of SEQ ID NO:11, SEQ ID NO:13, and the complement of SEQ ID NO:13, and modifying one or more nucleotides in the sequence to another nucleotide, deleting one or more nucleotides in the sequence, or adding one or more nucleotides to the sequence. In certain embodiments, the modifications are introduced by a method selected from the group consisting of error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, in vivo mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, ligation reassembly, GSSM and any combination thereof.

In a sixteenth aspect, the invention provides a computer readable medium having stored thereon a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, the complement of SEQ ID NO:1, SEQ ID NO:3, the complement of SEQ ID NO:3, SEQ ID NO:5, the complement of SEQ ID NO:5, SEQ ID NO:7, the complement of SEQ ID NO:7, SEQ ID NO:9, the complement of SEQ ID NO:9, SEQ ID NO:11, the complement of SEQ ID NO:11, SEQ ID NO:13, the complement of SEQ ID NO:13, and sequences substantially identical thereto.

In a seventeenth aspect, the invention provides a computer readable medium having stored thereon a nucleic acid sequence selected from the group consisting of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, and sequences substantially identical thereto.

In an eighteenth aspect, the invention provides a computer system. The computer system comprises a processor and a data storage device wherein said data storage device has stored thereon a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, the complement of SEQ ID NO:1, SEQ ID NO:3, the complement of SEQ ID NO:3, SEQ ID NO:5, the complement of SEQ ID NO:5, SEQ ID NO:7, the complement of SEQ ID NO:7, SEQ ID NO:9, the complement of SEQ ID NO:9, SEQ ID NO:11, the complement of SEQ ID NO:11, SEQ ID NO:13, the complement of SEQ ID NO:13, and sequences substantially identical thereto.

In a nineteenth aspect, the invention provides a computer system comprising a processor and a data storage device, wherein said data storage device has stored thereon a nucleic acid sequence selected from the group consisting of a polypeptide sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, and sequences substantially identical thereto.

In certain embodiments of the eighteenth and nineteenth aspects of the invention, the computer system further comprises a sequence comparison algorithm and a data storage device having at least one reference sequence stored thereon. In an

embodiment thereof, the sequence comparison algorithm comprises a computer program which indicates polymorphisms. In other certain embodiments the eighteenth and nineteenth aspects of the invention, the computer system further comprising an identifier which identifies features in the sequence stored therein.

In a twentieth aspect, the invention provides a method for comparing a first sequence to a reference sequence. The method comprises reading the first sequence and the reference sequence through use of a computer program which compares sequences, and determining differences between the first sequence and the reference sequence with the computer program. The first sequence in this method is a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, the complement of SEQ ID NO:1, SEQ ID NO:3, the complement of SEQ ID NO:3, SEQ ID NO:5, the complement of SEQ ID NO:5, SEQ ID NO:7, the complement of SEQ ID NO:7, SEQ ID NO:9, the complement of SEQ ID NO:9, SEQ ID NO:11, the complement of SEQ ID NO:11, SEQ ID NO:13, the complement of SEQ ID NO:13, and sequences substantially identical thereto.

In a twenty-first aspect, the invention provides a method for comparing a first sequence to a reference sequence. The method comprises reading the first sequence and the reference sequence through use of a computer program which compares sequences, and determining differences between the first sequence and the reference sequence with the computer program. With this method, the first sequence is a polypeptide sequence has an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, and sequences substantially identical thereto.

In certain embodiments of the twentieth and twenty-first aspects, differences identified between the first sequence and the reference sequence comprises identifying polymorphisms.

In a twenty-second aspect, the invention provides a method for identifying a feature in a sequence. The method comprises reading the sequence through the use of a computer program which identifies features in sequences; and identifying features in

the sequences with the computer program. For this method, a sequence is a nucleic acid sequence having an amino acid sequence selected from the group consisting of SEQ ID NO:1, the complement of SEQ ID NO:1, SEQ ID NO:3, the complement of SEQ ID NO:3, SEQ ID NO:5, the complement of SEQ ID NO:5, SEQ ID NO:7, the complement of SEQ ID NO:7, SEQ ID NO:9, the complement of SEQ ID NO:9, SEQ ID NO:11, the complement of SEQ ID NO:11, SEQ ID NO:13, the complement of SEQ ID NO:13, and sequences substantially identical thereto.

In a twenty-third aspect, the invention provides a method for identifying a feature in a sequence. The method comprises reading the sequence through the use of a computer program which identifies features in sequences, and identifying features in the sequences with the computer program. Sequences utilized in this method include a polypeptide sequence having the amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14, and sequences substantially identical thereto.

In a twenty-fifth aspect, the invention provides a method of making a polypeptide having a sequence selected from the group consisting of in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14, and sequences substantially identical thereto. The method includes introducing a nucleic acid encoding the polypeptide into a host cell, wherein the nucleic acid is operably linked to a promoter, and culturing the host cell under conditions that allow expression of the nucleic acid.

In a twenty-sixth aspect, the invention provides a method of making a polypeptide having at least 10 amino acids of a sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14, and sequences substantially identical thereto. The method includes introducing a nucleic acid encoding the polypeptide into a host cell, wherein the nucleic acid is operably linked to a promoter, and culturing the host cell under conditions that allow expression of the nucleic acid.

In a twenty-seventh aspect, the invention provides a method to identity a phytate sequence comprising analyzing an amino acid sequence for the occurrence of a first region consisting of RHGVRXaaPT and a second region consisting of WPXaaWPV, wherein the first and second region are separated by 13 amino acids, wherein Xaa can be any amino acid. In various embodiments thereof, the first and the second region are separated by 10, 11, 12, 14, 15, and 16 amino acids.

These and other aspects of the present invention should be apparent to those skilled in the art from the teachings herein.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

Figure 1 is a block diagram of a computer system.

Figure 2 is a flow diagram illustrating one embodiment of a process for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database.

Figure 3 is a flow diagram illustrating one embodiment of a process in a computer for determining whether two sequences are homologous.

Figure 4 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence.

Figure 5A is a representation of the nucleotide sequence of the Y. pestis phytase sequence identified by BLAST analaysis.

Figure 5B is a representation of the deduced amino acid sequences of the *Y. pestis* phytase sequence identified by BLAST analaysis.

Figure 5C is a representation of the nucleotide sequence of the corrected *Y. pestis* phytase sequence identified by BLAST analaysis.

Figure 5D is a representation of the deduced amino acid sequences of the corrected *Y. pestis* phytase sequence identified by BLAST analaysis.

Figure 5E is a representation of the nucleotide sequence of the 953-6 phytase sequence.

Figure 5F is a representation of the deduced amino acid sequences for the 953-6 phytase sequence.

Figure 5G is a representation of the nucleotide sequence of the Rhizobium phytase sequence.

Figure 5H is a representation of the deduced amino acid sequences for the Rhizobium phytase sequence.

Figure 5I is a representation of the nucleotide sequence of the 954-2 phytase sequence.

Figure 5J is a representation of the deduced amino acid sequences for the 954-2 phytase sequence.

Figure 5K is a representation of the nucleotide sequence of the *Y. pestis* expressed phytase sequence.

Figure 5L is a representation of the deduced amino acid sequences for the *Y. pestis* expressed phytase sequence.

Figure 5M is a representation of the nucleotide sequence of the *Y. pestis* consensus phytase sequence.

Figure 5N is a representation of the deduced amino acid sequences for the *Y. pestis* consensus phytase sequence.

Figure 6 shows an amino acid alignment of the phytases of the invention (SEQ ID Nos:4, 6, 8, 10, and 14).

Figure 7A presents a pictorial demonstrating results of a phytase overlay assay performed on isolates from the re-transformation of SEQ ID NO:11 phytase plasmid DNA.

Figure 7B presents a pictorial demonstrating results of a phytase overlay assay on Ed1#21, a control isolate lacking a lot of phytase activity, and Ed1#22 (SEQ ID NO:11), an isolate displaying phytase activity.

DETAILED DESCRIPTION OF THE INVENTION

The invention relates to phytase polypeptides and polynucleotides encoding them as well as methods of use of the polynucleotides and polypeptides. As used herein, the terminology "phytase" encompasses enzymes having phytase activity, for example, enzymes capable of catalyzing the degredation of phytate.

The phytases and polynucleotides encoding the phytases of the invention are useful in a number of processes, methods, and compositions. For example, as discussed above, a phytase can be used in animal feed, and feed supplements as well as in treatments to degrade or remove excess phytate from the environment or a sample. Other uses will be apparent to those of skill in the art based upon the teachings provided herein, including those discussed above.

The present invention provides purified recombinant phytase enzymes, shown in Figure 5-6. Additionally, the present invention provides isolated nucleic acid molecules (polynucleotides) which encode for the mature enzyme having an amino acid sequences as set forth in Figure 1.

The phytase molecules of the instant invention (particularly the recombinant enzyme and the polynucleotides that encode it) are novel with respect to their structures and with respect to their origin. Additionally, the instant phytase molecules have novel activity. For example, using an assay (as described in Food Chemicals Codex, 4th Ed.) the activity of the instant phytase enzyme was demonstrated to be far superior in comparison to a fungal (*Aspergillus*) phytase control.

The present invention provides purified a recombinant enzyme that catalyzes the hydrolysis of phytate to inositol and free phosphate with release of minerals from the phytic acid complex. An exemplary purified enzyme has a sequence as shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14.

Definitions

The phrases "nucleic acid" or "nucleic acid sequence" as used herein refer to an oligonucleotide, nucleotide, polynucleotide, or to a fragment of any of these, to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent a sense or antisense strand, peptide nucleic acid (PNA), or to any DNA-like or RNA-like material, natural or synthetic in origin. In one embodiment, a "nucleic acid sequence" of the invention includes, for example, a sequence encoding a polypeptide as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8,

SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14 and variants thereof. In another embodiment, a "nucleic acid sequence" of the invention includes, for example, a sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, sequences complemetary thereto, fragments of the foregoing sequences and variants thereof.

A "coding sequence" or a "nucleotide sequence encoding" a particular polypeptide or protein, is a nucleic acid sequence which is transcribed and translated into a polypeptide or protein when placed under the control of appropriate regulatory sequences.

The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as, where applicable, intervening sequences (introns) between individual coding segments (exons).

"Amino acid" or "amino acid sequence" as used herein refer to an oligopeptide, peptide, polypeptide, or protein sequence, or to a fragment, portion, or subunit of any of these, and to naturally occurring or synthetic molecules. In one embodiment, an "amino acid sequence" or "polypeptide sequence" of the invention includes, for example, a sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, fragments of the foregoing sequences and variants thereof. In another embodiment, an "amino acid sequence" of the invention includes, for example, a sequence encoded by a polynucleotide having a sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, sequences complemetary thereto, fragments of the foregoing sequences and variants thereof.

The term "polypeptide" as used herein, refers to amino acids joined to each other by peptide bonds or modified peptide bonds, *i.e.*, peptide isosteres, and may contain modified amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art.

Modifications can occur anywhere in the polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also a given polypeptide may have many types of modifications. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of a phosphytidylinositol, cross-linking cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristolyation, oxidation, pergylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, and transfer-RNA mediated addition of amino acids to protein such as arginylation. (See Proteins – Structure and Molecular Properties 2nd Ed., T.E. Creighton, W.H. Freeman and Company, New York (1993); Posttranslational Covalent Modification of Proteins, B.C. Johnson, Ed., Academic Press, New York, pp. 1-12 (1983)).

As used herein, the term "isolated" means that the material is removed from its original environment (*e.g.*, the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual nucleic acids obtained from a library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these clones could not be obtained directly either from the library or from total human DNA. The purified nucleic acids of the invention have been purified from the remainder of the genomic DNA in the organism by at least 10^4-10^6 fold. However, the

term "purified" also includes nucleic acids which have been purified from the remainder of the genomic DNA or from other sequences in a library or other environment by at least one order of magnitude, typically two or three orders, and more typically four or five orders of magnitude.

As used herein, the term "recombinant" means that the nucleic acid is adjacent to "backbone" nucleic acid to which it is not adjacent in its natural environment. Additionally, to be "enriched" the nucleic acids will represent 5% or more of the number of nucleic acid inserts in a population of nucleic acid backbone molecules. Backbone molecules according to the invention include nucleic acids such as expression vectors, self-replicating nucleic acids, viruses, integrating nucleic acids, and other vectors or nucleic acids used to maintain or manipulate a nucleic acid insert of interest. Typically, the enriched nucleic acids represent 15% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. More typically, the enriched nucleic acids represent 50% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. In a one embodiment, the enriched nucleic acids represent 90% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules.

"Recombinant" polypeptides or proteins refer to polypeptides or proteins produced by recombinant DNA techniques; *i.e.*, produced from cells transformed by an exogenous DNA construct encoding the desired polypeptide or protein. "Synthetic" polypeptides or protein are those prepared by chemical synthesis. Solid-phase chemical peptide synthesis methods can also be used to synthesize the polypeptide or fragments of the invention. Such method have been known in the art since the early 1960's (Merrifield, R. B., *J. Am. Chem. Soc.*, 85:2149-2154, 1963) (See also Stewart, J. M. and Young, J. D., Solid Phase Peptide Synthesis, 2 ed., Pierce Chemical Co., Rockford, Ill., pp. 11-12)) and have recently been employed in commercially available laboratory peptide design and synthesis kits (Cambridge Research Biochemicals). Such commercially available laboratory kits have generally utilized the teachings of H. M. Geysen *et al, Proc. Natl. Acad. Sci., USA*, 81:3998 (1984) and provide for synthesizing peptides upon the tips of a multitude of "rods" or "pins" all of which are connected to a single plate. When such a system is utilized, a plate of rods or pins is inverted and inserted into a second plate of

corresponding wells or reservoirs, which contain solutions for attaching or anchoring an appropriate amino acid to the pin's or rod's tips. By repeating such a process step, *i.e.*, inverting and inserting the rod's and pin's tips into appropriate solutions, amino acids are built into desired peptides. In addition, a number of available FMOC peptide synthesis systems are available. For example, assembly of a polypeptide or fragment can be carried out on a solid support using an Applied Biosystems, Inc. Model 431A automated peptide synthesizer. Such equipment provides ready access to the peptides of the invention, either by direct synthesis or by synthesis of a series of fragments that can be coupled using other known techniques.

A promoter sequence is "operably linked to" a coding sequence when RNA polymerase which initiates transcription at the promoter will transcribe the coding sequence into mRNA.

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described herein are known in the art and will be apparent to the ordinarily skilled artisan.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 μg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 μl of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the gel electrophoresis may be performed to isolate the desired fragment.

"Oligonucleotide" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

The phrase "substantially identical" in the context of two nucleic acid sequences or polypeptides, refers to two or more sequences that have at least 60%, 70%, 80%, and in some aspects 90-95% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using one of the known sequence comparison algorithms or by visual inspection. Typically, the substantial identity exists over a region of at least about 100 residues, and most commonly the sequences are substantially identical over at least about 150-200 residues. In some embodiments, the sequences are substantially identical over the entire length of the coding regions.

The term "about" is used herein to mean "approximately," or "roughly," or "around," or "in the region of." When the term "about" is used in conjunction with a numerical range, it modifies that range by extending the boundaries above and below the numerical values set forth. In general, the term "about" is used herein to modify a numerical value above and below the stated value by a variance of 20 percent.

Additionally a "substantially identical" amino acid sequence is a sequence that differs from a reference sequence by one or more conservative or non-conservative amino acid substitutions, deletions, or insertions, particularly when such a substitution occurs at a site that is not the active site of the molecule, and provided that the polypeptide essentially retains its functional properties. A conservative amino acid substitution, for example, substitutes one amino acid for another of the same class (*e.g.*, substitution of one hydrophobic amino acid, such as isoleucin, valine, leucine, or methionine, for another, or substitution of one polar amino acid for another, such as substitution of arginine for lysine, glutamic acid for aspartic acid or glutamine for asparagine). One or more amino acids can be deleted, for example, from a phytase polypeptide, resulting in modification of the structure of the polypeptide, without

significantly altering its biological activity. For example, amino- or carboxyl-terminal amino acids that are not required for phytase biological activity can be removed. Modified polypeptide sequences of the invention can be assayed for phytase biological activity by any number of methods, including contacting the modified polypeptide sequence with an phytase substrate and determining whether the modified polypeptide decreases the amount of specific substrate in the assay or increases the bioproducts of the enzymatic reaction of a functional phytase polypeptide with the substrate.

"Fragments" as used herein are a portion of a naturally occurring or recombinant protein which can exist in at least two different conformations. Fragments can have the same or substantially the same amino acid sequence as the naturally occurring protein. "Substantially the same" means that an amino acid sequence is largely, but not entirely, the same, but retains at least one functional activity of the sequence to which it is related. In general two amino acid sequences are "substantially the same" or "substantially homologous" if they are at least about 70, but more typically about 85% or more identical. Fragments which have different three dimensional structures as the naturally occurring protein are also included. An example of this, is a "pro-form" molecule, such as a low activity proprotein that can be modified by cleavage to produce a mature enzyme with significantly higher activity.

"Hybridization" refers to the process by which a nucleic acid strand joins with a complementary strand through base pairing. Hybridization reactions can be sensitive and selective so that a particular sequence of interest can be identified even in samples in which it is present at low concentrations. Suitably stringent conditions can be defined by, for example, the concentrations of salt or formamide in the prehybridization and hybridization solutions, or by the hybridization temperature, and are well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

For example, hybridization under high stringency conditions could occur in about 50% formamide at about 37°C to 42°C. Hybridization could occur under reduced stringency conditions in about 35% to 25% formamide at about 30°C to 35°C. In particular, hybridization could occur under high stringency conditions at 42°C in 50%

formamide, 5X SSPE, 0.3% SDS, and 200 ng/ml sheared and denatured salmon sperm DNA. Hybridization could occur under reduced stringency conditions as described above, but in 35% formamide at a reduced temperature of 35°C. The temperature range corresponding to a particular level of stringency can be further narrowed by calculating the purine to pyrimidine ratio of the nucleic acid of interest and adjusting the temperature accordingly. Variations on the above ranges and conditions are well known in the art.

The term "variant" refers to polynucleotides or polypeptides of the invention modified at one or more base pairs, codons, introns, exons, or amino acid residues (respectively) yet still retain the biological activity of an phytase of the invention. Variants can be produced by any number of means including methods such as, for example, error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, in vivo mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, ligation reassembly, GSSM and any combination thereof.

In one aspect, a non-stochastic method termed synthetic ligation reassembly (SLR), that is somewhat related to stochastic shuffling, save that the nucleic acid building blocks are not shuffled or concatenated or chimerized randomly, but rather are assembled non-stochastically can be used to create variants.

The SLR method does not depend on the presence of a high level of homology between polynucleotides to be shuffled. The invention can be used to non-stochastically generate libraries (or sets) of progeny molecules comprised of over 10^{100} different chimeras. Conceivably, SLR can even be used to generate libraries comprised of over 10^{1000} different progeny chimeras.

Thus, in one aspect, the invention provides a non-stochastic method of producing a set of finalized chimeric nucleic acid molecules having an overall assembly order that is chosen by design, which method is comprised of the steps of generating by design a plurality of specific nucleic acid building blocks having serviceable mutually

compatible ligatable ends, and assembling these nucleic acid building blocks, such that a designed overall assembly order is achieved.

The mutually compatible ligatable ends of the nucleic acid building blocks to be assembled are considered to be "serviceable" for this type of ordered assembly if they enable the building blocks to be coupled in predetermined orders. Thus, in one aspect, the overall assembly order in which the nucleic acid building blocks can be coupled is specified by the design of the ligatable ends and, if more than one assembly step is to be used, then the overall assembly order in which the nucleic acid building blocks can be coupled is also specified by the sequential order of the assembly step(s). In a one embodiment of the invention, the annealed building pieces are treated with an enzyme, such as a ligase (*e.g.*, T4 DNA ligase) to achieve covalent bonding of the building pieces.

In a another embodiment, the design of nucleic acid building blocks is obtained upon analysis of the sequences of a set of progenitor nucleic acid templates that serve as a basis for producing a progeny set of finalized chimeric nucleic acid molecules. These progenitor nucleic acid templates thus serve as a source of sequence information that aids in the design of the nucleic acid building blocks that are to be mutagenized, *i.e.* chimerized or shuffled.

In one exemplification, the invention provides for the chimerization of a family of related genes and their encoded family of related products. In a particular exemplification, the encoded products are enzymes. Enzymes and polypeptides for use in the invention can be mutagenized in accordance with the methods described herein.

Thus according to one aspect of the invention, the sequences of a plurality of progenitor nucleic acid templates are aligned in order to select one or more demarcation points, which demarcation points can be located at an area of homology. The demarcation points can be used to delineate the boundaries of nucleic acid building blocks to be generated. Thus, the demarcation points identified and selected in the progenitor molecules serve as potential chimerization points in the assembly of the progeny molecules.

Typically a serviceable demarcation point is an area of homology (comprised of at least one homologous nucleotide base) shared by at least two progenitor templates, but the demarcation point can be an area of homology that is shared by at least half of the progenitor templates, at least two thirds of the progenitor templates, at least three fourths of the progenitor templates, and preferably at almost all of the progenitor templates. Even more preferably still a serviceable demarcation point is an area of homology that is shared by all of the progenitor templates.

In a one embodiment, the ligation reassembly process is performed exhaustively in order to generate an exhaustive library. In other words, all possible ordered combinations of the nucleic acid building blocks are represented in the set of finalized chimeric nucleic acid molecules. At the same time, the assembly order (*i.e.* the order of assembly of each building block in the 5' to 3 sequence of each finalized chimeric nucleic acid) in each combination is by design (or non-stochastic). Because of the non-stochastic nature of the method, the possibility of unwanted side products is greatly reduced.

In another embodiment, the method provides that, the ligation reassembly process is performed systematically, for example in order to generate a systematically compartmentalized library, with compartments that can be screened systematically, *e.g.*, one by one. In other words the invention provides that, through the selective and judicious use of specific nucleic acid building blocks, coupled with the selective and judicious use of sequentially stepped assembly reactions, an experimental design can be achieved where specific sets of progeny products are made in each of several reaction vessels. This allows a systematic examination and screening procedure to be performed. Thus, it allows a potentially very large number of progeny molecules to be examined systematically in smaller groups.

Because of its ability to perform chimerizations in a manner that is highly flexible yet exhaustive and systematic as well, particularly when there is a low level of homology among the progenitor molecules, the instant invention provides for the generation of a library (or set) comprised of a large number of progeny molecules.

Because of the non-stochastic nature of the instant ligation reassembly invention, the progeny molecules generated preferably comprise a library of finalized chimeric nucleic acid molecules having an overall assembly order that is chosen by design. In a particularly embodiment, such a generated library is comprised of greater than 10^3 to greater than 10^{1000} different progeny molecular species.

In one aspect, a set of finalized chimeric nucleic acid molecules, produced as described is comprised of a polynucleotide encoding a polypeptide. According to one embodiment, this polynucleotide is a gene, which may be a man-made gene. According to another embodiment, this polynucleotide is a gene pathway, which may be a man-made gene pathway. The invention provides that one or more man-made genes generated by the invention may be incorporated into a man-made gene pathway, such as pathway operable in a eukaryotic organism (including a plant).

In another exemplifaction, the synthetic nature of the step in which the building blocks are generated allows the design and introduction of nucleotides (*e.g.*, one or more nucleotides, which may be, for example, codons or introns or regulatory sequences) that can later be optionally removed in an *in vitro* process (*e.g.*, by mutageneis) or in an *in vivo* process (*e.g.*, by utilizing the gene splicing ability of a host organism). It is appreciated that in many instances the introduction of these nucleotides may also be desirable for many other reasons in addition to the potential benefit of creating a serviceable demarcation point.

Thus, according to another embodiment, the invention provides that a nucleic acid building block can be used to introduce an intron. Thus, the invention provides that functional introns may be introduced into a man-made gene of the invention. The invention also provides that functional introns may be introduced into a man-made gene pathway of the invention. Accordingly, the invention provides for the generation of a chimeric polynucleotide that is a man-made gene containing one (or more) artificially introduced intron(s).

Accordingly, the invention also provides for the generation of a chimeric polynucleotide that is a man-made gene pathway containing one (or more) artificially

introduced intron(s). Preferably, the artificially introduced intron(s) are functional in one or more host cells for gene splicing much in the way that naturally-occurring introns serve functionally in gene splicing. The invention provides a process of producing man-made intron-containing polynucleotides to be introduced into host organisms for recombination and/or splicing.

A man-made genes produced using the invention can also serve as a substrate for recombination with another nucleic acid. Likewise, a man-made gene pathway produced using the invention can also serve as a substrate for recombination with another nucleic acid. In a preferred instance, the recombination is facilitated by, or occurs at, areas of homology between the man-made intron-containing gene and a nucleic acid with serves as a recombination partner. In a particularly preferred instance, the recombination partner may also be a nucleic acid generated by the invention, including a man-made gene or a man-made gene pathway. Recombination may be facilitated by or may occur at areas of homology that exist at the one (or more) artificially introduced intron(s) in the man-made gene.

The synthetic ligation reassembly method of the invention utilizes a plurality of nucleic acid building blocks, each of which preferably has two ligatable ends. The two ligatable ends on each nucleic acid building block may be two blunt ends (*i.e.* each having an overhang of zero nucleotides), or preferably one blunt end and one overhang, or more preferably still two overhangs.

A useful overhang for this purpose may be a 3' overhang or a 5' overhang. Thus, a nucleic acid building block may have a 3' overhang or alternatively a 5' overhang or alternatively two 3' overhangs or alternatively two 5' overhangs. The overall order in which the nucleic acid building blocks are assembled to form a finalized chimeric nucleic acid molecule is determined by purposeful experimental design and is not random.

According to one preferred embodiment, a nucleic acid building block is generated by chemical synthesis of two single-stranded nucleic acids (also referred to as

single-stranded oligos) and contacting them so as to allow them to anneal to form a double-stranded nucleic acid building block.

A double-stranded nucleic acid building block can be of variable size. The sizes of these building blocks can be small or large. Preferred sizes for building block range from 1 base pair (not including any overhangs) to 100,000 base pairs (not including any overhangs). Other preferred size ranges are also provided, which have lower limits of from 1 bp to 10,000 bp (including every integer value in between), and upper limits of from 2 bp to 100, 000 bp (including every integer value in between).

Many methods exist by which a double-stranded nucleic acid building block can be generated that is serviceable for the invention; and these are known in the art and can be readily performed by the skilled artisan.

According to one embodiment, a double-stranded nucleic acid building block is generated by first generating two single stranded nucleic acids and allowing them to anneal to form a double-stranded nucleic acid building block. The two strands of a double-stranded nucleic acid building block may be complementary at every nucleotide apart from any that form an overhang; thus containing no mismatches, apart from any overhang(s). According to another embodiment, the two strands of a double-stranded nucleic acid building block are complementary at fewer than every nucleotide apart from any that form an overhang. Thus, according to this embodiment, a double-stranded nucleic acid building block can be used to introduce codon degeneracy. Preferably the codon degeneracy is introduced using the site-saturation mutagenesis described herein, using one or more N,N,G/T cassettes or alternatively using one or more N,N,N cassettes.

The *in vivo* recombination method of the invention can be performed blindly on a pool of unknown hybrids or alleles of a specific polynucleotide or sequence. However, it is not necessary to know the actual DNA or RNA sequence of the specific polynucleotide.

The approach of using recombination within a mixed population of genes can be useful for the generation of any useful proteins, for example, interleukin I, antibodies, tPA and growth hormone. This approach may be used to generate proteins having altered specificity or activity. The approach may also be useful for the generation of hybrid nucleic acid sequences, for example, promoter regions, introns, exons, enhancer sequences, 31 untranslated regions or 51 untranslated regions of genes. Thus this approach may be used to generate genes having increased rates of expression. This approach may also be useful in the study of repetitive DNA sequences. Finally, this approach may be useful to mutate ribozymes or aptamers.

In one aspect variants of the polynucleotides and polypeptides described herein are obtained by the use of repeated cycles of reductive reassortment, recombination and selection which allow for the directed molecular evolution of highly complex linear sequences, such as DNA, RNA or proteins thorough recombination.

In vivo shuffling of molecules is useful in providing variants and can be performed utilizing the natural property of cells to recombine multimers. While recombination *in vivo* has provided the major natural route to molecular diversity, genetic recombination remains a relatively complex process that involves 1) the recognition of homologies; 2) strand cleavage, strand invasion, and metabolic steps leading to the production of recombinant chiasma; and finally 3) the resolution of chiasma into discrete recombined molecules. The formation of the chiasma requires the recognition of homologous sequences.

In a another embodiment, the invention includes a method for producing a hybrid polynucleotide from at least a first polynucleotide and a second polynucleotide. The invention can be used to produce a hybrid polynucleotide by introducing at least a first polynucleotide and a second polynucleotide which share at least one region of partial sequence homology (*e.g.,* SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, and combinations thereof) into a suitable host cell. The regions of partial sequence homology promote processes which result in sequence reorganization producing a hybrid polynucleotide. The term "hybrid polynucleotide", as used herein, is any nucleotide sequence which results from

the method of the present invention and contains sequence from at least two original polynucleotide sequences. Such hybrid polynucleotides can result from intermolecular recombination events which promote sequence integration between DNA molecules. In addition, such hybrid polynucleotides can result from intramolecular reductive reassortment processes which utilize repeated sequences to alter a nucleotide sequence within a DNA molecule.

The invention provides a means for generating hybrid polynucleotides which may encode biologically active hybrid polypeptides (*e.g.*, a hybrid phytase). In one aspect, the original polynucleotides encode biologically active polypeptides. The method of the invention produces new hybrid polypeptides by utilizing cellular processes which integrate the sequence of the original polynucleotides such that the resulting hybrid polynucleotide encodes a polypeptide demonstrating activities derived from the original biologically active polypeptides. For example, the original polynucleotides may encode a particular enzyme from different microorganisms. An enzyme encoded by a first polynucleotide from one organism or variant may, for example, function effectively under a particular environmental condition, *e.g.*, high salinity. An enzyme encoded by a second polynucleotide from a different organism or variant may function effectively under a different environmental condition, such as extremely high temperatures. A hybrid polynucleotide containing sequences from the first and second original polynucleotides may encode an enzyme which exhibits characteristics of both enzymes encoded by the original polynucleotides. Thus, the enzyme encoded by the hybrid polynucleotide may function effectively under environmental conditions shared by each of the enzymes encoded by the first and second polynucleotides, *e.g.*, high salinity and extreme temperatures.

Enzymes encoded by original polynucleotides include, but are not limited to, hydrolases and phytases. A hybrid polypeptide resulting from the method of the invention may exhibit specialized enzyme activity not displayed in the original enzymes. For example, following recombination and/or reductive reassortment of polynucleotides encoding hydrolase activities, the resulting hybrid polypeptide encoded by a hybrid polynucleotide can be screened for specialized hydrolase activities obtained from each of the original enzymes, *i.e.*, the type of bond on which the hydrolase acts

and the temperature at which the hydrolase functions. Thus, for example, the hydrolase may be screened to ascertain those chemical functionalities which distinguish the hybrid hydrolase from the original hydrolyases, such as: (a) amide (peptide bonds), *i.e.*, proteases; (b) ester bonds, *i.e.*, esterases and lipases; (c) acetals, *i.e.*, glycosidases and, for example, the temperature, pH or salt concentration at which the hybrid polypeptide functions.

Sources of the original polynucleotides may be isolated from individual organisms ("isolates"), collections of organisms that have been grown in defined media ("enrichment cultures"), or, uncultivated organisms ("environmental samples"). The use of a culture-independent approach to derive polynucleotides encoding novel bioactivities from environmental samples is most preferable since it allows one to access untapped resources of biodiversity.

"Environmental libraries" are generated from environmental samples and represent the collective genomes of naturally occurring organisms archived in cloning vectors that can be propagated in suitable prokaryotic hosts. Because the cloned DNA is initially extracted directly from environmental samples, the libraries are not limited to the small fraction of prokaryotes that can be grown in pure culture. Additionally, a normalization of the environmental DNA present in these samples could allow more equal representation of the DNA from all of the species present in the original sample. This can dramatically increase the efficiency of finding interesting genes from minor constituents of the sample which may be under-represented by several orders of magnitude compared to the dominant species.

For example, gene libraries generated from one or more uncultivated microorganisms are screened for an activity of interest. Potential pathways encoding bioactive molecules of interest are first captured in prokaryotic cells in the form of gene expression libraries. Polynucleotides encoding activities of interest are isolated from such libraries and introduced into a host cell. The host cell is grown under conditions which promote recombination and/or reductive reassortment creating potentially active biomolecules with novel or enhanced activities.

The microorganisms from which the polynucleotide may be prepared include prokaryotic microorganisms, such as *Xanthobacter*, *Eubacteria* and *Archaebacteria*, and lower eukaryotic microorganisms such as fungi, some algae and protozoa. Polynucleotides may be isolated from environmental samples in which case the nucleic acid may be recovered without culturing of an organism or recovered from one or more cultured organisms. In one aspect, such microorganisms may be extremophiles, such as hyperthermophiles, psychrophiles, psychrotrophs, halophiles, barophiles and acidophiles. Polynucleotides encoding enzymes isolated from extremophilic microorganisms are particularly preferred. Such enzymes may function at temperatures above 100°C in terrestrial hot springs and deep sea thermal vents, at temperatures below 0°C in arctic waters, in the saturated salt environment of the Dead Sea, at pH values around 0 in coal deposits and geothermal sulfur-rich springs, or at pH values greater than 11 in sewage sludge. For example, several esterases and lipases cloned and expressed from extremophilic organisms show high activity throughout a wide range of temperatures and pHs.

Polynucleotides selected and isolated as hereinabove described are introduced into a suitable host cell. A suitable host cell is any cell which is capable of promoting recombination and/or reductive reassortment. The selected polynucleotides are preferably already in a vector which includes appropriate control sequences. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or preferably, the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis *et al.*, 1986).

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as *E. coli*, *Streptomyces*, *Salmonella typhimurium*; fungal cells, such as yeast; insect cells such as *Drosophila S2* and *Spodoptera Sf9*; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; and plant cells. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

The majority of bioactive compounds currently in use are derived from soil microorganisms. Many microbes inhabiting soils and other complex ecological communities produce a variety of compounds that increase their ability to survive and proliferate. These compounds are generally thought to be nonessential for growth of the organism and are synthesized with the aid of genes involved in intermediary metabolism hence their name – "secondary metabolites". Secondary metabolites are generally the products of complex biosynthetic pathways and are usually derived from common cellular precursors. Secondary metabolites that influence the growth or survival of other organisms are known as "bioactive" compounds and serve as key components of the chemical defense arsenal of both micro- and macro-organisms. Humans have exploited these compounds for use as antibiotics, antiinfectives and other bioactive compounds with activity against a broad range of prokaryotic and eukaryotic pathogens. Approximately 6,000 bioactive compounds of microbial origin have been characterized, with more than 60% produced by the gram positive soil bacteria of the genus Streptomyces. (Barnes et al., Proc.Nat. Acad. Sci. U.S.A., 91, 1994).

Hybridization screening using high density filters or biopanning has proven an efficient approach to detect homologues of pathways containing genes of interest to discover novel bioactive molecules that may have no known counterparts. Once a polynucleotide of interest is enriched in a library of clones it may be desirable to screen for an activity. For example, it may be desirable to screen for the expression of small molecule ring structures or "backbones". Because the genes encoding these polycyclic structures can often be expressed in *E. coli*, the small molecule backbone can be manufactured, even if in an inactive form. Bioactivity is conferred upon transferring the molecule or pathway to an appropriate host that expresses the requisite glycosylation and methylation genes that can modify or "decorate" the structure to its active form. Thus, even if inactive ring compounds, recombinantly expressed in *E. coli* are detected to identify clones which are then shuttled to a metabolically rich host, such as *Streptomyces* (e.g., *Streptomyces diversae* or *venezuelae*) for subsequent production of the bioactive molecule. It should be understood that *E. coli* can produce active small molecules and in certain instances it may be desirable to shuttle clones to a metabolically rich host for "decoration" of the structure, but not required. The use of

high throughput robotic systems allows the screening of hundreds of thousands of clones in multiplexed arrays in microtiter dishes.

With particular references to various mammalian cell culture systems that can be employed to express recombinant protein, examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in "SV40-transformed simian cells support the replication of early SV40 mutants" (Gluzman, 1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Host cells containing the polynucleotides of interest can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan. The clones which are identified as having the specified enzyme activity may then be sequenced to identify the polynucleotide sequence encoding an enzyme having the enhanced activity.

The enzymes and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity. The phytase polypeptide of the invention can be obtained using any of several standard methods. For example, phytase polypeptides can be produced in a standard recombinant expression system (as described herein), chemically synthesized (although somewhat limited to small phytase peptide fragments), or purified from organisms in which they are naturally expressed. Useful recombinant expression methods include mammalian hosts, microbial hosts, and plant hosts.

The recombinant expression of the instant phytase molecules may be achieved in combination with one or more additional molecules such as, for example, other enzymes. This approach is useful for producing combination products, such as a plant or plant part that contains the instant phytase molecules as well as one or more additional molecules – preferably the phytase molecules and the additional molecules are used in a combination treatment. The resulting recombinantly expresssed molecules may be used in homogenized and/or purified form or alternatively in relatively unpurified form (*e.g.* as consumable plant parts that are useful when admixed with other foodstuffs for catalyzing the degredation of phytate).

In sum, in a non-limiting embodiment, the present invention provides a recombinant enzyme expressed in a host. In another non-limiting embodiment, the present invention provides a substantially pure phytase enzyme. Thus, an enzyme of the present invention may be a recombinant enzyme, a natural enzyme, or a synthetic enzyme, preferably a recombinant enzyme.

In a particular embodiment, the present invention provides for the expression of phytase in transgenic plants or plant organs and methods for the production thereof. DNA expression constructs are provided for the transformation of plants with a gene encoding phytase under the control of regulatory sequences which are capable of directing the expression of phytase. These regulatory sequences include sequences capable of directing transcription in plants, either constitutively, or in stage and/or tissue specific manners.

The manner of expression depends, in part, on the use of the plant or parts thereof. The transgenic plants and plant organs provided by the present invention may be applied to a variety of industrial processes either directly, *e.g.* in animal feeds or alternatively, the expressed phytase may be extracted and if desired, purified before application. Alternatively, the recombinant host plant or plant part may be used directly. In a particular aspect, the present invention provides methods of catalyzing phytate-hydrolyzing reactions using seeds containing enhanced amounts of phytase. The method involves contacting transgenic, non-wild type seeds, preferably in a ground or chewed form, with phytate-containing substrate and allowing the enzymes in the

seeds to increase the rate of reaction. By directly adding the seeds to a phytate-containing substrate, the invention provides a solution to the expensive and problematic process of extracting and purifying the enzyme. In a particular – but by no means limiting – exemplification, the present invention also provides methods of treatment whereby an organism lacking a sufficient supply of an enzyme is administered the enzyme in the form of seeds containing enhanced amounts of the enzyme. In a preferred embodiment, the timing of the administration of the enzyme to an organism is coordinated with the consumption of a phytate-containing foodstuff.

The expression of phytase in plants can be achieved by a variety of means. Specifically, for example, technologies are available for transforming a large number of plant species, including dicotyledonous species (*e.g.* tobacco, potato, tomato, Petunia, Brassica). Additionally, for example, strategies for the expression of foreign genes in plants are available. Additionally still, regulatory sequences from plant genes have been identified that are serviceable for the construction of chimeric genes that can be functionally expressed in plants and in plant cells (*e.g.* Klee *et al.*, 1987; Clark *et al.*, 1990; Smith *et al.*, 1990).

The introduction of gene constucts into plants can be achieved using several technologies including transformation with *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*. Non-limiting examples of plant tissues that can be transformed thusly include protoplasts, microspores or pollen, and explants such as leaves, stems, roots, hypocotyls, and cotyls. Furthermore, DNA can be introduced directly into protoplasts and plant cells or tissues by microinjection, electriporation, particle bombardment, and direct DNA uptake.

Proteins may be produced in plants by a variety of expression systems. For instance, the use of a constitutive promoter such as the 35S promoter of Cauliflower Mosaic Virus (Guilley *et al.*, 1982) is serviceable for the accumulation of the expressed protein in virtually all organs of the transgenic plant. Alternatively, the use of promoters that are highly tissue-specific and/or stage-specific are serviceable for this invention (Higgins, 1984; Shotwell, 1989) in order to bias expression towards desired tissues and/or towards a desired stage of development. Further details relevant to the

expression in plants of the phytase molecules of the instant invention are disclosed, for example, in U.S. Patent No. 5,770,413 (Van Ooijen *et al.*) and U.S. Patent No. 5,593,963 (Van Ooijen *et al.*), although these reference do not teach the inventive molecules of the instant application and instead teach the use of fungal phytases.

In sum, it is relevant to this invention that a variety of means can be used to achieve the recombinant expression of phytase in a transgenic plant or plant part. Such a transgenic plants and plant parts are serviceable as sources of recombinantly expressed phytase, which can be added directly to phytate-containing sources. Alternatively, the recombinant plant-expressed phytase can be extracted away from the plant source and, if desired, purified prior to contacting the phytase substrate.

Within the context of the present invention, plants to be selected include, but are not limited to crops producing edible flowers such as cauliflower (*Brassica oleracea*), artichoke (*Cynara scolymus*), fruits such as apple (Malus, *e.g. domesticus*), banana (Musa, *e.g. acuminata*), berries (such as the currant, Ribes, *e.g. rubrum*), cherries (such as the sweet cherry, Prunus, *e.g. avium*), cucumber (Cucumis, *e.g. sativus*), grape (Vitis, *e.g. vinifera*), lemon (*Citrus limon*), melon (*Cucumis melo*), nuts (such as the walnut, Juglans, *e.g. regia*; peanut, Arachis hypogeae), orange (Citrus, *e.g. maxima*), peach (Prunus, *e.g. persica*), pear (Pyra, *e.g. communis*), plum (Prunus, *e.g. domestica*), strawberry (Fragaria, *e.g. moschata*), tomato (Lycopersicon, *e.g. esculentum*), leafs, such as alfalfa (Medicago, *e.g. sativa*), cabbages (*e.g. Brassica oleracea*), endive (Cichoreum, *e.g. endivia*), leek (Allium, *e.g. porrum*), lettuce (Lactuca, *e.g. sativa*), spinach (Spinacia, *e.g. oleraceae*), tobacco (Nicotiana, *e.g. tabacum*), roots, such as arrowroot (Maranta, *e.g. arundinacea*), beet (Beta, *e.g. vulgaris*), carrot (Daucus, *e.g. carota*), cassava (Manihot, *e.g. esculenta*), turnip (Brassica, *e.g. rapa*), radish (Raphanus, *e.g. sativus*), yam (Dioscorea, *e.g. esculenta*), sweet potato (*Ipomoea batatas*) and seeds, such as bean (Phaseolus, *e.g. vulgaris*), pea (Pisum, *e.g. sativum*), soybean (Glycin, *e.g. max*), wheat (Triticum, *e.g. aestivum*), barley (Hordeum, *e.g. vulgare*), corn (Zea, *e.g. mays*), rice (Oryza, *e.g. sativa*), rapeseed (*Brassica napus*), millet (Panicum L.), sunflower (*Helianthus annus*), oats (*Avena sativa*), tubers, such as kohlrabi (Brassica, *e.g. oleraceae*), potato (Solanum, *e.g. tuberosum*) and the like.

It is understood that additional plant as well as non-plant expression systems can be used within the context of this invention. The choice of the plant species is primarily determined by the intended use of the plant or parts thereof and the amenability of the plant species to transformation.

Several techniques are available for the introduction of the expression construct containing the phytase-encoding DNA sequence into the target plants. Such techniques include but are not limited to transformation of protoplasts using the calcium/polyethylene glycol method, electroporation and microinjection or (coated) particle bombardment (Potrykus, 1990). In addition to these so-called direct DNA transformation methods, transformation systems involving vectors are widely available, such as viral vectors (*e.g.* from the Cauliflower Mosaic Cirus (CaMV) and bacterial vectors (*e.g.* from the genus Agrobacterium) (Potrykus, 1990). After selection and/or screening, the protoplasts, cells or plant parts that have been transformed can be regenerated into whole plants, using methods known in the art (Horsch *et al.*, 1985). The choice of the transformation and/or regeneration techniques is not critical for this invention.

For dicots, a preferred embodiment of the present invention uses the principle of the binary vector system (Hoekema *et al.*, 1983; EP 0120516 Schilperoort *et al.*) in which Agrobacterium strains are used which contain a vir plasmid with the virulence genes and a compatible plasmid containing the gene construct to be transferred. This vector can replicate in both *E. coli* and in Agrobacterium, and is derived from the binary vector Bin19 (Bevan, 1984) which is altered in details that are not relevant for this invention. The binary vectors as used in this example contain between the left- and right-border sequences of the T-DNA, an identical NPTII-gene coding for kanamycin resistance (Bevan, 1984) and a multiple cloning site to clone in the required gene constructs.

The transformation and regeneration of monocotyledonous crops is not a standard procedure. However, recent scientific progress shows that in principle monocots are amenable to transformation and that fertile transgenic plants can be regenerated from transformed cells. The development of reproducible tissue culture

systems for these crops, together with the powerful methods for introduction of genetic material into plant cells has facilitated transformation. Presently the methods of choice for transformation of monocots are microprojectile bombardment of explants or suspension cells, and direct DNA uptake or electroporation of protoplasts. For example, transgenic rice plants have been successfully obtained using the bacterial hph gene, encoding hygromycin resistance, as a selection marker. The gene was introduced by electroporation (Shimamoto *et al.*, 1993). Transgenic maize plants have been obtained by introducing the *Streptomyces hygroscopicus* bar gene, which encodes phosphinothricin acetyltransferase (an enzyme which inactivates the herbicide phosphinothricin), into embryogenic cells of a maize suspension culture by microparticle bombardment (Gordon-Kamm *et al.*, 1990). The introduction of genetic material into aleurone protoplasts of other monocot crops such as wheat and barley has been reported (Lee *et al.*, 1989). Wheat plants have been regenerated from embryogenic suspension culture by selecting only the aged compact and nodular embryogenic callus tissues for the establishment of the embryogenic suspension cultures (Vasil *et al.*, 1972: Vasil *et al.*, 1974). The combination with transformation systems for these crops enables the application of the present invention to monocots. These methods may also be applied for the transformation and regeneration of dicots.

Expression of the phytase construct involves such details as transcription of the gene by plant polymerases, translation of mRNA, etc. that are known to persons skilled in the art of recombinant DNA techniques. Only details relevant for the proper understanding of this invention are discussed below. Regulatory sequences which are known or are found to cause expression of phytase may be used in the present invention. The choice of the regulatory sequences used depends on the target crop and/or target organ of interest. Such regulatory sequences may be obtained from plants or plant viruses, or may be chemically synthesized. Such regulatory sequences are promoters active in directing transcription in plants, either constitutively or stage and/or tissue specific, depending on the use of the plant or parts thereof. These promoters include, but are not limited to promoters showing constitutive expression, such as the 35S promoter of Cauliflower Mosaic Virus (CaMV) (Guilley *et al.*, 1982), those for leaf-specific expression, such as the promoter of the ribulose bisphosphate carboxylase small subunit gene (Coruzzi *et al.*, 1984), those for root-specific expression, such as the

promoter from the glutamin synthase gene (Tingey *et al.*, 1987), those for seed-specific expression, such as the cruciferin A promoter from *Brassica napus* (Ryan *et al.*, 1989), those for tuber-specific expression, such as the class-I patatin promoter from potato (Koster-Topfer *et al.*, 1989; Wenzler *et al.*, 1989) or those for fruit-specific expression, such as the polygalacturonase (PG) promoter from tomato (Bird *et al.*, 1988).

Other regulatory sequences such as terminator sequences and polyadenylation signals include any such sequence functioning as such in plants, the choice of which is within the level of the skilled artisan. An example of such sequences is the 3′ flanking region of the nopaline synthase (nos) gene of *Agrobacterium tumefaciens* (Bevan, *supra*). The regulatory sequences may also include enhancer sequences, such as found in the 35S promoter of CaMV, and mRNA stabilizing sequences such as the leader sequence of Alfalfa Mosaic Cirus (AlMV) RNA4 (Brederode *et al.*, 1980) or any other sequences functioning in a like manner.

The phytase should be expressed in an environment that allows for stability of the expressed protein. The choice of cellular compartments, such as cytosol, endoplasmic reticulum, vacuole, protein body or periplasmic space can be used in the present invention to create such a stable environment, depending on the biophysical parameters of the phytase. Such parameters include, but are not limited to pH-optimum, sensitivity to proteases or sensitivity to the molarity of the preferred compartment.

To obtain expression in the cytoplasm of the cell, the expressed enzyme should not contain a secretory signal peptide or any other target sequence. For expression in chloroplasts and mitochondria the expressed enzyme should contain specific so-called transit peptide for import into these organelles. Targeting sequences that can be attached to the enzyme of interest in order to achieve this are known (Smeekens *et al.*, 1990; van den Broeck *et al.*, 1985; Wolter *et al.*, 1988). If the activity of the enzyme is desired in the vacuoles a secretory signal peptide has to be present, as well as a specific targeting sequence that directs the enzyme to these vacuoles (Tague *et al.*, 1990). The same is true for the protein bodies in seeds. The DNA sequence encoding the enzyme

of interest should be modified in such a way that the enzyme can exert its action at the desired location in the cell.

To achieve extracellular expression of the phytase, the expression construct of the present invention utilizes a secretory signal sequence. Although signal sequences which are homologous (native) to the plant host species are preferred, heterologous signal sequences, *i.e.* those originating from other plant species or of microbial origin, may be used as well. Such signal sequences are known to those skilled in the art. Appropriate signal sequences which may be used within the context of the present invention are disclosed in Blobel *et al.*, 1979; Von Heijne, 1986; Garcia *et al.*, 1987; Sijmons *et al.*, 1990; Ng *et al.*, 1994; and Powers *et al.*, 1996).

All parts of the relevant DNA constructs (promoters, regulatory-, secretory-, stabilizing-, targeting-, or termination sequences) of the present invention may be modified, if desired, to affect their control characteristics using methods known to those skilled in the art. It is pointed out that plants containing phytase obtained via the present invention may be used to obtain plants or plant organs with yet higher phytase levels. For example, it may be possible to obtain such plants or plant organs by the use of somoclonal variation techniques or by cross breeding techniques. Such techniques are well known to those skilled in the art.

In one embodiment, the instant invention provides a method (and products thereof) of achieving a highly efficient overexpression system for phytase and other molecules. In a preferred embodiment, the instant invention provides a method (and products thereof) of achieving a highly efficient overexpression system for phytase and pH 2.5 acid phosphatase in Trichoderma. This system results in enzyme compositions that have particular utility in the animal feed industry.

Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes EP 0659215 (W0 9403612 A1) (Nevalainen *et al.*), although these reference do not teach the inventive molecules of the instant application.

In another aspect, methods can be used to generate novel polynucleotides encoding biochemical pathways from one or more operons or gene clusters or portions

thereof. For example, bacteria and many eukaryotes have a coordinated mechanism for regulating genes whose products are involved in related processes. The genes are clustered, in structures referred to as "gene clusters," on a single chromosome or immediately adjacent to one another and are transcribed together under the control of a single regulatory sequence, including a single promoter which initiates transcription of the entire cluster. Thus, a gene cluster is a group of adjacent genes that are either identical or related, usually as to their function. An example of a biochemical pathway encoded by gene clusters are polyketides. Polyketides are molecules which are an extremely rich source of bioactivities, including antibiotics (such as tetracyclines and erythromycin), anti-cancer agents (daunomycin), immunosuppressants (FK506 and rapamycin), and veterinary products (monensin). Many polyketides (produced by polyketide synthases) are valuable as therapeutic agents. Polyketide synthases are multifunctional enzymes that catalyze the biosynthesis of an enormous variety of carbon chains differing in length and patterns of functionality and cyclization. Polyketide synthase genes fall into gene clusters and at least one type (designated type I) of polyketide synthases have large size genes and enzymes, complicating genetic manipulation and *in vitro* studies of these genes/proteins.

Gene cluster DNA can be isolated from different organisms and ligated into vectors, particularly vectors containing expression regulatory sequences which can control and regulate the production of a detectable protein or protein-related array activity from the ligated gene clusters. Use of vectors which have an exceptionally large capacity for exogenous DNA introduction are particularly appropriate for use with such gene clusters and are described by way of example herein to include the f-factor (or fertility factor) of *E. coli*. This f-factor of *E. coli* is a plasmid which affects high-frequency transfer of itself during conjugation and is ideal to achieve and stably propagate large DNA fragments, such as gene clusters from mixed microbial samples. Once ligated into an appropriate vector, two or more vectors containing different phytase gene clusters can be introduced into a suitable host cell. Regions of partial sequence homology shared by the gene clusters will promote processes which result in sequence reorganization resulting in a hybrid gene cluster. The novel hybrid gene cluster can then be screened for enhanced activities not found in the original gene clusters.

Therefore, in a one embodiment, the invention relates to a method for producing a biologically active hybrid polypeptide and screening such a polypeptide for enhanced activity by:

1) introducing at least a first polynucleotide in operable linkage and a second polynucleotide in operable linkage, said at least first polynucleotide and second polynucleotide sharing at least one region of partial sequence homology, into a suitable host cell;
2) growing the host cell under conditions which promote sequence reorganization resulting in a hybrid polynucleotide in operable linkage;
3) expressing a hybrid polypeptide encoded by the hybrid polynucleotide;
4) screening the hybrid polypeptide under conditions which promote identification of enhanced biological activity; and
5) isolating the a polynucleotide encoding the hybrid polypeptide.

Methods for screening for various enzyme activities are known to those of skill in the art and are discussed throughout the present specification. Such methods may be employed when isolating the polypeptides and polynucleotides of the invention.

As representative examples of expression vectors which may be used there may be mentioned viral particles, baculovirus, phage, plasmids, phagemids, cosmids, fosmids, bacterial artificial chromosomes, viral DNA (*e.g.*, vaccinia, adenovirus, foul pox virus, pseudorabies and derivatives of SV40), P1-based artificial chromosomes, yeast plasmids, yeast artificial chromosomes, and any other vectors specific for specific hosts of interest (such as bacillus, aspergillus and yeast). Thus, for example, the DNA may be included in any one of a variety of expression vectors for expressing a polypeptide. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences. Large numbers of suitable vectors are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example; Bacterial: pQE vectors (Qiagen), pBluescript plasmids, pNH vectors, (lambda-ZAP vectors (Stratagene); ptrc99a, pKK223-3, pDR540, pRIT2T (Pharmacia); Eukaryotic: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, pSVLSV40 (Pharmacia). However, any other plasmid or other vector may be used so long as they are replicable and viable

in the host. Low copy number or high copy number vectors may be employed with the present invention.

A preferred type of vector for use in the present invention contains an f-factor origin replication. The f-factor (or fertility factor) in *E. coli* is a plasmid which effects high frequency transfer of itself during conjugation and less frequent transfer of the bacterial chromosome itself. A particularly preferred embodiment is to use cloning vectors, referred to as "fosmids" or bacterial artificial chromosome (BAC) vectors. These are derived from *E. coli* f-factor which is able to stably integrate large segments of genomic DNA. When integrated with DNA from a mixed uncultured environmental sample, this makes it possible to achieve large genomic fragments in the form of a stable "environmental DNA library."

Another type of vector for use in the present invention is a cosmid vector. Cosmid vectors were originally designed to clone and propagate large segments of genomic DNA. Cloning into cosmid vectors is described in detail in "Molecular Cloning: A laboratory Manual" (Sambrook *et al.*, 1989).

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct RNA synthesis. Particular named bacterial promoters include *lacI*, *lacZ*, *T3*, *T7*, *gpt*, *lambda* P_R, P_L and *trp*. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression. Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or tetracycline or ampicillin resistance in *E. coli*.

In vivo reassortment is focused on "inter-molecular" processes collectively referred to as "recombination" which in bacteria, is generally viewed as a "RecA-dependent" phenomenon. The invention can rely on recombination processes of a host cell to recombine and re-assort sequences, or the cells' ability to mediate reductive processes to decrease the complexity of quasi-repeated sequences in the cell by deletion. This process of "reductive reassortment" occurs by an "intra-molecular", RecA-independent process.

Therefore, in another aspect of the invention, variant polynucleotides can be generated by the process of reductive reassortment. The method involves the generation of constructs containing consecutive sequences (original encoding sequences), their insertion into an appropriate vector, and their subsequent introduction into an appropriate host cell. The reassortment of the individual molecular identities occurs by combinatorial processes between the consecutive sequences in the construct possessing regions of homology, or between quasi-repeated units. The reassortment process recombines and/or reduces the complexity and extent of the repeated sequences, and results in the production of novel molecular species. Various treatments may be applied to enhance the rate of reassortment. These could include treatment with ultra-violet light, or DNA damaging chemicals, and/or the use of host cell lines displaying enhanced levels of "genetic instability". Thus the reassortment process may involve homologous recombination or the natural property of quasi-repeated sequences to direct their own evolution.

Repeated or "quasi-repeated" sequences play a role in genetic instability. In the present invention, "quasi-repeats" are repeats that are not restricted to their original unit structure. Quasi-repeated units can be presented as an array of sequences in a construct; consecutive units of similar sequences. Once ligated, the junctions between the consecutive sequences become essentially invisible and the quasi-repetitive nature of the resulting construct is now continuous at the molecular level. The deletion process the cell performs to reduce the complexity of the resulting construct operates between the quasi-repeated sequences. The quasi-repeated units provide a practically limitless repertoire of templates upon which slippage events can occur. The constructs containing the quasi-repeats thus effectively provide sufficient molecular elasticity that

deletion (and potentially insertion) events can occur virtually anywhere within the quasi-repetitive units.

When the quasi-repeated sequences are all ligated in the same orientation, for instance head to tail or vice versa, the cell cannot distinguish individual units. Consequently, the reductive process can occur throughout the sequences. In contrast, when for example, the units are presented head to head, rather than head to tail, the inversion delineates the endpoints of the adjacent unit so that deletion formation will favor the loss of discrete units. Thus, it is preferable with the present method that the sequences are in the same orientation. Random orientation of quasi-repeated sequences will result in the loss of reassortment efficiency, while consistent orientation of the sequences will offer the highest efficiency. However, while having fewer of the contiguous sequences in the same orientation decreases the efficiency, it can still provide sufficient elasticity for the effective recovery of novel molecules. Constructs can be made with the quasi-repeated sequences in the same orientation to allow higher efficiency.

Sequences can be assembled in a head to tail orientation using any of a variety of methods, including the following:

a) Primers that include a poly-A head and poly-T tail which when made single-stranded provide orientation can be utilized. This is accomplished by having the first few bases of the primers made from RNA and hence easily removed RNAseH.

b) Primers that include unique restriction cleavage sites can be utilized. Multiple sites, a battery of unique sequences, and repeated synthesis and ligation steps would be required.

c) The inner few bases of the primer can be thiolated and an exonuclease used to produce properly tailed molecules.

The recovery of the re-assorted sequences relies on the identification of cloning vectors with a reduced RI. The re-assorted encoding sequences can then be recovered by amplification. The products are re-cloned and expressed. The recovery of cloning vectors with reduced RI can be effected by:

1) The use of vectors only stably maintained when the construct is reduced in complexity;

2) The physical recovery of shortened vectors by physical procedures. In this case, the cloning vector is recovered using standard plasmid isolation procedures and size fractionated on either an agarose gel, or column with a low molecular weight cut off utilizing standard procedures;

3) The recovery of vectors containing interrupted genes which can be selected when insert size decreases; and

4) The use of direct selection techniques with an expression vector and the appropriate selection.

Encoding sequences (for example, genes) from related organisms may demonstrate a high degree of homology and encode quite diverse protein products. These types of sequences are particularly useful in the present invention as quasi-repeats. However, while the examples illustrated below demonstrate the reassortment of nearly identical original encoding sequences (quasi-repeats), this process is not limited to such nearly identical repeats.

The following example demonstrates a method of the invention. Encoding nucleic acid sequences (quasi-repeats) derived from three unique species are depicted. Each sequence encodes a protein with a distinct set of properties. Each of the sequences differs by a single or a few base pairs at a unique position in the sequence which are designated "A", "B" and "C". The quasi-repeated sequences are separately or collectively amplified and ligated into random assemblies such that all possible permutations and combinations are available in the population of ligated molecules. The number of quasi-repeat units can be controlled by the assembly conditions. The average number of quasi-repeated units in a construct is defined as the repetitive index (RI).

Once formed, the constructs may or may not be size fractionated on an agarose gel according to published protocols, inserted into a cloning vector, and transfected into an appropriate host cell. The cells are then propagated and "reductive reassortment" is effected. The rate of the reductive reassortment process may be stimulated by the

introduction of DNA damage if desired. Whether the reduction in RI is mediated by deletion formation between repeated sequences by an "intra-molecular" mechanism, or mediated by recombination-like events through "inter-molecular" mechanisms is immaterial. The end result is a reassortment of the molecules into all possible combinations.

Optionally, the method comprises the additional step of screening the library members of the shuffled pool to identify individual shuffled library members having the ability to bind or otherwise interact, or catalyze a particular reaction (*e.g.*, such as catalyzing the hydrolysis of a phytate).

The polypeptides that are identified from such libraries can be used for therapeutic, diagnostic, research and related purposes (*e.g.*, catalysts, solutes for increasing osmolarity of an aqueous solution, and the like), and/or can be subjected to one or more additional cycles of shuffling and/or selection.

In another aspect, prior to or during recombination or reassortment, polynucleotides of the invention or polynucleotides generated by the method described herein can be subjected to agents or processes which promote the introduction of mutations into the original polynucleotides. The introduction of such mutations would increase the diversity of resulting hybrid polynucleotides and polypeptides encoded therefrom. The agents or processes which promote mutagenesis can include, but are not limited to: (+)-CC-1065, or a synthetic analog such as (+)-CC-1065-(N3-Adenine, see Sun and Hurley, 1992); an N-acelylated or deacetylated 4'-fluro-4-aminobiphenyl adduct capable of inhibiting DNA synthesis (see, for example, van de Poll *et al.*, 1992); or a N-acetylated or deacetylated 4-aminobiphenyl adduct capable of inhibiting DNA synthesis (see also, van de Poll *et al.*, 1992, pp. 751-758); trivalent chromium, a trivalent chromium salt, a polycyclic aromatic hydrocarbon ("PAH") DNA adduct capable of inhibiting DNA replication, such as 7-bromomethyl-benz[*a*]anthracene ("BMA"), tris(2,3-dibromopropyl)phosphate ("Tris-BP"), 1,2-dibromo-3-chloropropane ("DBCP"), 2-bromoacrolein (2BA), benzo[*a*]pyrene-7,8-dihydrodiol-9-10-epoxide ("BPDE"), a platinum(II) halogen salt, N-hydroxy-2-amino-3-methylimidazo[4,5-*f*]-quinoline ("N-hydroxy-IQ"), and N-hydroxy-2-amino-1-methyl-6-phenylimidazo[4,5-

f]-pyridine ("N-hydroxy-PhIP"). Especially preferred means for slowing or halting PCR amplification consist of UV light (+)-CC-1065 and (+)-CC-1065-(N3-Adenine). Particularly encompassed means are DNA adducts or polynucleotides comprising the DNA adducts from the polynucleotides or polynucleotides pool, which can be released or removed by a process including heating the solution comprising the polynucleotides prior to further processing.

In another aspect the invention is directed to a method of producing recombinant proteins having biological activity by treating a sample comprising double-stranded template polynucleotides encoding a wild-type protein under conditions according to the invention which provide for the production of hybrid or re-assorted polynucleotides.

The invention also provides for the use of proprietary codon primers (containing a degenerate N,N,G/T sequence) to introduce point mutations into a polynucleotide, so as to generate a set of progeny polypeptides in which a full range of single amino acid substitutions is represented at each amino acid position (gene site saturated mutagenesis (GSSM)). The oligos used are comprised contiguously of a first homologous sequence, a degenerate N,N,G/T sequence, and preferably but not necessarily a second homologous sequence. The downstream progeny translational products from the use of such oligos include all possible amino acid changes at each amino acid site along the polypeptide, because the degeneracy of the N,N,G/T sequence includes codons for all 20 amino acids.

In one aspect, one such degenerate oligo (comprised of one degenerate N,N,G/T cassette) is used for subjecting each original codon in a parental polynucleotide template to a full range of codon substitutions. In another aspect, at least two degenerate N,N,G/T cassettes are used – either in the same oligo or not, for subjecting at least two original codons in a parental polynucleotide template to a full range of codon substitutions. Thus, more than one N,N,G/T sequence can be contained in one oligo to introduce amino acid mutations at more than one site. This plurality of N,N,G/T sequences can be directly contiguous, or separated by one or more additional nucleotide sequence(s). In another aspect, oligos serviceable for introducing additions

and deletions can be used either alone or in combination with the codons containing an N,N,G/T sequence, to introduce any combination or permutation of amino acid additions, deletions, and/or substitutions.

In a particular exemplification, it is possible to simultaneously mutagenize two or more contiguous amino acid positions using an oligo that contains contiguous N,N,G/T triplets, *i.e.* a degenerate $(N,N,G/T)_n$ sequence.

In another aspect, the present invention provides for the use of degenerate cassettes having less degeneracy than the N,N,G/T sequence. For example, it may be desirable in some instances to use (*e.g.* in an oligo) a degenerate triplet sequence comprised of only one N, where said N can be in the first second or third position of the triplet. Any other bases including any combinations and permutations thereof can be used in the remaining two positions of the triplet. Alternatively, it may be desirable in some instances to use (*e.g.*, in an oligo) a degenerate N,N,N triplet sequence, or an N,N, G/C triplet sequence.

It is appreciated, however, that the use of a degenerate triplet (such as N,N,G/T or an N,N, G/C triplet sequence) as disclosed in the instant invention is advantageous for several reasons. In one aspect, this invention provides a means to systematically and fairly easily generate the substitution of the full range of possible amino acids (for a total of 20 amino acids) into each and every amino acid position in a polypeptide. Thus, for a 100 amino acid polypeptide, the invention provides a way to systematically and fairly easily generate 2000 distinct species (*i.e.*, 20 possible amino acids per position times 100 amino acid positions). It is appreciated that there is provided, through the use of an oligo containing a degenerate N,N,G/T or an N,N, G/C triplet sequence, 32 individual sequences that code for 20 possible amino acids. Thus, in a reaction vessel in which a parental polynucleotide sequence is subjected to saturation mutagenesis using one such oligo, there are generated 32 distinct progeny polynucleotides encoding 20 distinct polypeptides. In contrast, the use of a non-degenerate oligo in site-directed mutagenesis leads to only one progeny polypeptide product per reaction vessel.

This invention also provides for the use of nondegenerate oligos, which can optionally be used in combination with degenerate primers disclosed. It is appreciated that in some situations, it is advantageous to use nondegenerate oligos to generate specific point mutations in a working polynucleotide. This provides a means to generate specific silent point mutations, point mutations leading to corresponding amino acid changes, and point mutations that cause the generation of stop codons and the corresponding expression of polypeptide fragments.

Thus, in one embodiment, each saturation mutagenesis reaction vessel contains polynucleotides encoding at least 20 progeny polypeptide molecules such that all 20 amino acids are represented at the one specific amino acid position corresponding to the codon position mutagenized in the parental polynucleotide. The 32-fold degenerate progeny polypeptides generated from each saturation mutagenesis reaction vessel can be subjected to clonal amplification (*e.g.*, cloned into a suitable *E. coli* host using an expression vector) and subjected to expression screening. When an individual progeny polypeptide is identified by screening to display a favorable change in property (when compared to the parental polypeptide), it can be sequenced to identify the correspondingly favorable amino acid substitution contained therein.

It is appreciated that upon mutagenizing each and every amino acid position in a parental polypeptide using saturation mutagenesis as disclosed herein, favorable amino acid changes may be identified at more than one amino acid position. One or more new progeny molecules can be generated that contain a combination of all or part of these favorable amino acid substitutions. For example, if 2 specific favorable amino acid changes are identified in each of 3 amino acid positions in a polypeptide, the permutations include 3 possibilities at each position (no change from the original amino acid, and each of two favorable changes) and 3 positions. Thus, there are 3 x 3 x 3 or 27 total possibilities, including 7 that were previously examined - 6 single point mutations (*i.e.*, 2 at each of three positions) and no change at any position.

In yet another aspect, site-saturation mutagenesis can be used together with shuffling, chimerization, recombination and other mutagenizing processes, along with screening. This invention provides for the use of any mutagenizing process(es),

including saturation mutagenesis, in an iterative manner. In one exemplification, the iterative use of any mutagenizing process(es) is used in combination with screening.

Thus, in a non-limiting exemplification, polynucleotides and polypeptides of the invention can be derived by saturation mutagenesis in combination with additional mutagenization processes, such as process where two or more related polynucleotides are introduced into a suitable host cell such that a hybrid polynucleotide is generated by recombination and reductive reassortment.

In addition to performing mutagenesis along the entire sequence of a gene, mutagenesis can be used to replace each of any number of bases in a polynucleotide sequence, wherein the number of bases to be mutagenized is preferably every integer from 15 to 100,000. Thus, instead of mutagenizing every position along a molecule, one can subject every or a discrete number of bases (preferably a subset totaling from 15 to 100,000) to mutagenesis. Preferably, a separate nucleotide is used for mutagenizing each position or group of positions along a polynucleotide sequence. A group of 3 positions to be mutagenized may be a codon. The mutations are preferably introduced using a mutagenic primer, containing a heterologous cassette, also referred to as a mutagenic cassette. Preferred cassettes can have from 1 to 500 bases. Each nucleotide position in such heterologous cassettes be N, A, C, G, T, A/C, A/G, A/T, C/G, C/T, G/T, C/G/T, A/G/T, A/C/T, A/C/G, or E, where E is any base that is not A, C, G, or T (E can be referred to as a designer oligo).

In a general sense, saturation mutagenesis is comprised of mutagenizing a complete set of mutagenic cassettes (wherein each cassette is preferably about 1-500 bases in length) in defined polynucleotide sequence to be mutagenized (wherein the sequence to be mutagenized is preferably from about 15 to 100,000 bases in length). Thus, a group of mutations (ranging from 1 to 100 mutations) is introduced into each cassette to be mutagenized. A grouping of mutations to be introduced into one cassette can be different or the same from a second grouping of mutations to be introduced into a second cassette during the application of one round of saturation mutagenesis. Such groupings are exemplified by deletions, additions, groupings of particular codons, and groupings of particular nucleotide cassettes.

Defined sequences to be mutagenized include a whole gene, pathway, cDNA, an entire open reading frame (ORF), and entire promoter, enhancer, repressor/transactivator, origin of replication, intron, operator, or any polynucleotide functional group. Generally, a "defined sequences" for this purpose may be any polynucleotide that a 15 base-polynucleotide sequence, and polynucleotide sequences of lengths between 15 bases and 15,000 bases (this invention specifically names every integer in between). Considerations in choosing groupings of codons include types of amino acids encoded by a degenerate mutagenic cassette.

In a particularly preferred exemplification a grouping of mutations that can be introduced into a mutagenic cassette, this invention specifically provides for degenerate codon substitutions (using degenerate oligos) that code for 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, and 20 amino acids at each position, and a library of polypeptides encoded thereby.

One aspect of the invention is an isolated nucleic acid comprising one of the sequences of sequences substantially identical thereto, sequences complementary thereto, or a fragment comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive bases of one of the sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13. The isolated, nucleic acids may comprise DNA, including cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. Alternatively, the isolated nucleic acids may comprise RNA.

As discussed in more detail below, the isolated nucleic acid sequences of the invention may be used to prepare one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14, and sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids of one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID

NO:10, SEQ ID NO:12 and SEQ ID NO:14, and sequences substantially identical thereto.

Accordingly, another aspect of the invention is an isolated nucleic acid sequence which encodes one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14 sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids of one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14. The coding sequences of these nucleic acids may be identical to one of the coding sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, or a fragment thereof, or may be different coding sequences which encode one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14, and sequences substantially identical thereto, and fragments having at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids of one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14 as a result of the redundancy or degeneracy of the genetic code. The genetic code is well known to those of skill in the art and can be obtained, for example, on page 214 of B. Lewin, Genes VI, Oxford University Press, 1997, the disclosure of which is incorporated herein by reference.

The isolated nucleic acid sequence which encodes one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14, and sequences substantially identical thereto, may include, but is not limited to only a coding sequence of one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, and sequences substantially identical thereto, and additional coding sequences, such as leader sequences or proprotein sequences and non-coding sequences, such as introns or non-coding sequences 5' and/or 3' of the coding sequence. Thus, as used herein, the term "polynucleotide encoding a polypeptide" encompasses a polynucleotide which includes only coding sequence for the polypeptide as well as a polynucleotide which includes additional coding and/or non-coding sequence.

Alternatively, the nucleic acid sequences of the invention may be mutagenized using conventional techniques, such as site directed mutagenesis, or other techniques familiar to those skilled in the art, to introduce silent changes into the polynucleotides of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, and sequences substantially identical thereto. As used herein, "silent changes" include, for example, changes which do not alter the amino acid sequence encoded by the polynucleotide. Such changes may be desirable in order to increase the level of the polypeptide produced by host cells containing a vector encoding the polypeptide by introducing codons or codon pairs which occur frequently in the host organism.

The invention also relates to polynucleotides which have nucleotide changes which result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptides of the invention (*e.g.*, SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14). Such nucleotide changes may be introduced using techniques such as site directed mutagenesis, random chemical mutagenesis, exonuclease III deletion, and other recombinant DNA techniques. Alternatively, such nucleotide changes may be naturally occurring allelic variants which are isolated by identifying nucleic acid sequences which specifically hybridize to probes comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive bases of one of the sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, and sequences substantially identical thereto, (or the sequences complementary thereto), under conditions of high, moderate, or low stringency as provided herein.

The isolated nucleic acids of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, sequences substantially identical thereto, complementary sequences, or a fragment comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive bases of one of the foregoing sequences, may also be used as probes to determine whether a biological sample, such as a soil sample, contains an organism having a nucleic acid sequence of the invention or an organism from which the nucleic acid was obtained. In such

procedures, a biological sample potentially harboring the organism from which the nucleic acid was isolated is obtained and nucleic acids are obtained from the sample. The nucleic acids are contacted with the probe under conditions which permit the probe to specifically hybridize to any complementary sequences which are present therein.

Where necessary, conditions which permit the probe to specifically hybridize to complementary sequences may be determined by placing the probe in contact with complementary sequences from samples known to contain the complementary sequence as well as control sequences which do not contain the complementary sequence. Hybridization conditions, such as the salt concentration of the hybridization buffer, the formamide concentration of the hybridization buffer, or the hybridization temperature, may be varied to identify conditions which allow the probe to hybridize specifically to complementary nucleic acids.

If the sample contains the organism from which the nucleic acid was isolated, specific hybridization of the probe is then detected. Hybridization may be detected by labeling the probe with a detectable agent such as a radioactive isotope, a fluorescent dye or an enzyme capable of catalyzing the formation of a detectable product.

Many methods for using the labeled probes to detect the presence of complementary nucleic acids in a sample are familiar to those skilled in the art. These include Southern Blots, Northern Blots, colony hybridization procedures, and dot blots. Protocols for each of these procedures are provided in Ausubel *et al.* Current Protocols in Molecular Biology, John Wiley 503 Sons, Inc. 1997 and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, 1989, the entire disclosures of which are incorporated herein by reference.

Alternatively, more than one probe (at least one of which is capable of specifically hybridizing to any complementary sequences which are present in the nucleic acid sample), may be used in an amplification reaction to determine whether the sample contains an organism containing a nucleic acid sequence of the invention (*e.g.*, an organism from which the nucleic acid was isolated). Typically, the probes comprise oligonucleotides. In one embodiment, the amplification reaction may comprise a PCR

reaction. PCR protocols are described in Ausubel and Sambrook, *supra*. Alternatively, the amplification may comprise a ligase chain reaction, 3SR, or strand displacement reaction. (See Barany, F., "The Ligase Chain Reaction in a PCR World," *PCR Methods and Applications* 1:5-16, 1991; E. Fahy *et al.*, "Self-sustained Sequence Replication (3SR): An Isothermal Transcription-based Amplification System Alternative to PCR", *PCR Methods and Applications* 1:25-33, 1991; and Walker G.T. *et al.*, "Strand Displacement Amplification-an Isothermal *in vitro* DNA Amplification Technique", *Nucleic Acid Research* 20:1691-1696, 1992, the disclosures of which are incorporated herein by reference in their entireties). In such procedures, the nucleic acids in the sample are contacted with the probes, the amplification reaction is performed, and any resulting amplification product is detected. The amplification product may be detected by performing gel electrophoresis on the reaction products and staining the gel with an intercalator such as ethidium bromide. Alternatively, one or more of the probes may be labeled with a radioactive isotope and the presence of a radioactive amplification product may be detected by autoradiography after gel electrophoresis.

Probes derived from sequences near the ends of a sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, and sequences substantially identical thereto, may also be used in chromosome walking procedures to identify clones containing genomic sequences located adjacent to the nucleic acid sequences as set forth above. Such methods allow the isolation of genes which encode additional proteins from the host organism.

An isolated nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, sequences substantially identical thereto, sequences complementary thereto, or a fragment comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive bases of one of the foregoing sequences may be used as probes to identify and isolate related nucleic acids. In some embodiments, the related nucleic acids may be cDNAs or genomic DNAs from organisms other than the one from which the nucleic acid was isolated. For example, the other organisms may be related organisms. In such procedures, a nucleic acid sample is contacted with the probe under conditions which permit the probe to specifically hybridize to related sequences.

Hybridization of the probe to nucleic acids from the related organism is then detected using any of the methods described above.

In nucleic acid hybridization reactions, the conditions used to achieve a particular level of stringency will vary, depending on the nature of the nucleic acids being hybridized. For example, the length, degree of complementarity, nucleotide sequence composition (*e.g.*, GC v. AT content), and nucleic acid type (*e.g.*, RNA v. DNA) of the hybridizing regions of the nucleic acids can be considered in selecting hybridization conditions. An additional consideration is whether one of the nucleic acids is immobilized, for example, on a filter.

Hybridization may be carried out under conditions of low stringency, moderate stringency or high stringency. As an example of nucleic acid hybridization, a polymer membrane containing immobilized denatured nucleic acids is first prehybridized for 30 minutes at 45°C in a solution consisting of 0.9 M NaCl, 50 mM NaH_2PO_4, pH 7.0, 5.0 mM Na_2EDTA, 0.5% SDS, 10X Denhardt's, and 0.5 mg/ml polyriboadenylic acid. Approximately 2 x 10^7 cpm (specific activity 4-9 x 10^8 cpm/ug) of ^{32}P end-labeled oligonucleotide probe are then added to the solution. After 12-16 hours of incubation, the membrane is washed for 30 minutes at room temperature in 1X SET (150 mM NaCl, 20 mM Tris hydrochloride, pH 7.8, 1 mM Na_2EDTA) containing 0.5% SDS, followed by a 30 minute wash in fresh 1X SET at Tm-10°C for the oligonucleotide probe. The membrane is then exposed to auto-radiographic film for detection of hybridization signals.

By varying the stringency of the hybridization conditions used to identify nucleic acids, such as cDNAs or genomic DNAs, which hybridize to the detectable probe, nucleic acids having different levels of homology to the probe can be identified and isolated. Stringency may be varied by conducting the hybridization at varying temperatures below the melting temperatures of the probes. The melting temperature, T_m, is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly complementary probe. Very stringent conditions are selected to be equal to or about 5°C lower than the T_m for a particular probe. The melting temperature of the probe may be calculated using the following formulas:

For probes between 14 and 70 nucleotides in length the melting temperature (T_m) is calculated using the formula: $T_m = 81.5 + 16.6(\log [Na+]) + 0.41(\text{fraction G+C}) - (600/N)$, where N is the length of the probe.

If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation: $T_m = 81.5 + 16.6(\log [Na+]) + 0.41(\text{fraction G+C}) - (0.63\% \text{ formamide}) - (600/N)$, where N is the length of the probe.

Prehybridization may be carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 µg denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 µg denatured fragmented salmon sperm DNA, 50% formamide. The formulas for SSC and Denhardt's solutions are listed in Sambrook *et al.*, *supra*.

Hybridization is conducted by adding the detectable probe to the prehybridization solutions listed above. Where the probe comprises double stranded DNA, it is denatured before addition to the hybridization solution. The filter is contacted with the hybridization solution for a sufficient period of time to allow the probe to hybridize to cDNAs or genomic DNAs containing sequences complementary thereto or homologous thereto. For probes over 200 nucleotides in length, the hybridization may be carried out at 15-25°C below the Tm. For shorter probes, such as oligonucleotide probes, the hybridization may be conducted at 5-10°C below the T_m. Typically, for hybridizations in 6X SSC, the hybridization is conducted at approximately 68°C. Usually, for hybridizations in 50% formamide containing solutions, the hybridization is conducted at approximately 42°C.

All of the foregoing hybridizations are considered to be under conditions of high stringency.

Following hybridization, the filter is washed to remove any non-specifically bound detectable probe. The stringency used to wash the filters can also be varied depending on the nature of the nucleic acids being hybridized, the length of the nucleic acids being hybridized, the degree of complementarity, the nucleotide sequence composition (*e.g.*, GC v. AT content), and the nucleic acid type (*e.g.*, RNA v. DNA).

Examples of progressively higher stringency condition washes are as follows: 2X SSC, 0.1% SDS at room temperature for 15 minutes (low stringency); 0.1X SSC, 0.5% SDS at room temperature for 30 minutes to 1 hour (moderate stringency); 0.1X SSC, 0.5% SDS for 15 to 30 minutes at between the hybridization temperature and 68°C (high stringency); and 0.15M NaCl for 15 minutes at 72°C (very high stringency). A final low stringency wash can be conducted in 0.1X SSC at room temperature. The examples above are merely illustrative of one set of conditions that can be used to wash filters. One of skill in the art would know that there are numerous recipes for different stringency washes. Some other examples are given below.

Nucleic acids which have hybridized to the probe are identified by autoradiography or other conventional techniques.

The above procedure may be modified to identify nucleic acids having decreasing levels of homology to the probe sequence. For example, to obtain nucleic acids of decreasing homology to the detectable probe, less stringent conditions may be used. For example, the hybridization temperature may be decreased in increments of 5°C from 68°C to 42°C in a hybridization buffer having a Na+ concentration of approximately 1 M. Following hybridization, the filter may be washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate" conditions above 50°C and "low" conditions below 50°C. A specific example of "moderate" hybridization conditions is when the above hybridization is conducted at 55°C. A specific example of "low stringency" hybridization conditions is when the above hybridization is conducted at 45°C.

Alternatively, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42°C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50°C. These conditions are considered to be "moderate" conditions above 25% formamide and "low" conditions below 25% formamide. A specific example of "moderate" hybridization conditions is when the above hybridization is conducted at 30% formamide. A specific

example of "low stringency" hybridization conditions is when the above hybridization is conducted at 10% formamide.

For example, the preceding methods may be used to isolate nucleic acids having a sequence with at least about 97%, at least 95%, at least 90%, at least 85%, at least 80%, or at least 70% homology to a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, sequences substantially identical thereto, or fragments comprising at least about 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive bases thereof, and the sequences complementary to any of the foregoing sequences. Homology may be measured using an alignment algorithm. For example, the homologous polynucleotides may have a coding sequence which is a naturally occurring allelic variant of one of the coding sequences described herein. Such allelic variants may have a substitution, deletion or addition of one or more nucleotides when compared to a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, or sequences complementary thereto.

Additionally, the above procedures may be used to isolate nucleic acids which encode polypeptides having at least about 99%, at least 95%, at least 90%, at least 85%, at least 80%, or at least 70% homology to a polypeptide having a sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof as determined using a sequence alignment algorithm (*e.g.*, such as the FASTA version 3.0t78 algorithm with the default parameters).

Another aspect of the invention is an isolated or purified polypeptide comprising a sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, sequences substantially identical thereto, or fragments comprising at least about 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof. As discussed above, such polypeptides may be obtained by inserting a nucleic acid encoding the polypeptide

into a vector such that the coding sequence is operably linked to a sequence capable of driving the expression of the encoded polypeptide in a suitable host cell. For example, the expression vector may comprise a promoter, a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

Promoters suitable for expressing the polypeptide or fragment thereof in bacteria include the *E. coli lac* or *trp* promoters, the *lacI* promoter, the *lacZ* promoter, the *T3* promoter, the *T7* promoter, the *gpt* promoter, the *lambda* P_R promoter, the *lambda* P_L promoter, promoters from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), and the acid phosphatase promoter. Fungal promoters include the α factor promoter. Eukaryotic promoters include the CMV immediate early promoter, the HSV thymidine kinase promoter, heat shock promoters, the early and late SV40 promoter, LTRs from retroviruses, and the mouse metallothionein-I promoter. Other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses may also be used.

Mammalian expression vectors may also comprise an origin of replication, any necessary ribosome binding sites, a polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. In some embodiments, DNA sequences derived from the SV40 splice and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Vectors for expressing the polypeptide or fragment thereof in eukaryotic cells may also contain enhancers to increase expression levels. Enhancers are cis-acting elements of DNA, usually from about 10 to about 300 bp in length that act on a promoter to increase its transcription. Examples include the SV40 enhancer on the late side of the replication origin bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and the adenovirus enhancers.

In addition, the expression vectors typically contain one or more selectable marker genes to permit selection of host cells containing the vector. Such selectable markers include genes encoding dihydrofolate reductase or genes conferring neomycin resistance for eukaryotic cell culture, genes conferring tetracycline or ampicillin resistance in *E. coli*, and the *S. cerevisiae TRP1* gene.

After the expression libraries have been generated, the additional step of "biopanning" such libraries prior to screening by cell sorting can be included. The "biopanning" procedure refers to a process for identifying clones having a specified biological activity by screening for sequence homology in a library of clones prepared by *(i)* selectively isolating target DNA, from DNA derived from at least one microorganism, by use of at least one probe DNA comprising at least a portion of a DNA sequence encoding an biological having the specified biological activity; and *(ii)* optionally transforming a host with isolated target DNA to produce a library of clones which are screened for the specified biological activity.

The probe DNA used for selectively isolating the target DNA of interest from the DNA derived from at least one microorganism can be a full-length coding region sequence or a partial coding region sequence of DNA for an enzyme of known activity. The original DNA library can be preferably probed using mixtures of probes comprising at least a portion of the DNA sequence encoding an enzyme having the specified enzyme activity. These probes or probe libraries are preferably single-stranded and the microbial DNA which is probed has preferably been converted into single-stranded form. The probes that are particularly suitable are those derived from DNA encoding enzymes having an activity similar or identical to the specified enzyme activity which is to be screened.

The probe DNA should be at least about 10 bases and preferably at least 15 bases. In one embodiment, the entire coding region may be employed as a probe. Conditions for the hybridization in which target DNA is selectively isolated by the use of at least one DNA probe will be designed to provide a hybridization stringency of at least about 50% sequence identity, more particularly a stringency providing for a sequence identity of at least about 70%.

In nucleic acid hybridization reactions, the conditions used to achieve a particular level of stringency will vary, depending on the nature of the nucleic acids being hybridized. For example, the length, degree of complementarity, nucleotide sequence composition (*e.g.*, GC v. AT content), and nucleic acid type (*e.g.*, RNA v. DNA) of the hybridizing regions of the nucleic acids can be considered in selecting hybridization conditions. An additional consideration is whether one of the nucleic acids is immobilized, for example, on a filter.

An example of progressively higher stringency conditions is as follows: 2X SSC/0.1% SDS at about room temperature (hybridization conditions); 0.2X SSC/0.1% SDS at about room temperature (low stringency conditions); 0.2X SSC/0.1% SDS at about 42°C (moderate stringency conditions); and 0.1X SSC at about 68°C (high stringency conditions). Washing can be carried out using only one of these conditions, *e.g.*, high stringency conditions, or each of the conditions can be used, *e.g.*, for 10-15 minutes each, in the order listed above, repeating any or all of the steps listed. However, as mentioned above, optimal conditions will vary, depending on the particular hybridization reaction involved, and can be determined empirically.

Hybridization techniques for probing a microbial DNA library to isolate target DNA of potential interest are well known in the art and any of those which are described in the literature are suitable for use herein, particularly those which use a solid phase-bound, directly or indirectly bound, probe DNA for ease in separation from the remainder of the DNA derived from the microorganisms.

Preferably the probe DNA is "labeled" with one partner of a specific binding pair (*i.e.* a ligand) and the other partner of the pair is bound to a solid matrix to provide ease of separation of target from its source. The ligand and specific binding partner can be selected from, in either orientation, the following: (1) an antigen or hapten and an antibody or specific binding fragment thereof; (2) biotin or iminobiotin and avidin or streptavidin; (3) a sugar and a lectin specific therefor; (4) an enzyme and an inhibitor therefor; (5) an apoenzyme and cofactor; (6) complementary homopolymeric oligonucleotides; and (7) a hormone and a receptor therefor. The solid phase is

preferably selected from: (1) a glass or polymeric surface; (2) a packed column of polymeric beads; and (3) magnetic or paramagnetic particles.

Further, it is optional but desirable to perform an amplification of the target DNA that has been isolated. In this embodiment the target DNA is separated from the probe DNA after isolation. It is then amplified before being used to transform hosts. The double stranded DNA selected to include as at least a portion thereof a predetermined DNA sequence can be rendered single-stranded, subjected to amplification and reannealed to provide amplified numbers of selected double-stranded DNA. Numerous amplification methodologies are now well known in the art.

The selected DNA is then used for preparing a library for screening by transforming a suitable organism. Hosts, particularly those specifically identified herein as preferred, are transformed by artificial introduction of the vectors containing the target DNA by inoculation under conditions conducive for such transformation. The resultant libraries of transformed clones are then screened for clones which display activity for the enzyme of interest.

Having prepared a multiplicity of clones from DNA selectively isolated from an organism, such clones are screened for a specific enzyme activity and to identify the clones having the specified enzyme characteristics.

The screening for enzyme activity may be effected on individual expression clones or may be initially effected on a mixture of expression clones to ascertain whether or not the mixture has one or more specified enzyme activities. If the mixture has a specified enzyme activity, then the individual clones may be rescreened utilizing a FACS machine for such enzyme activity or for a more specific activity. Alternatively, encapsulation techniques such as gel microdroplets, may be employed to localize multiple clones in one location to be screened on a FACS machine for positive expressing clones within the group of clones which can then be broken out into individual clones to be screened again on a FACS machine to identify positive individual clones. Thus, for example, if a clone mixture has hydrolase activity, then the individual clones may be recovered and screened utilizing a FACS machine to

determine which of such clones has hydrolase activity. As used herein, "small insert library" means a gene library containing clones with random small size nucleic acid inserts of up to approximately 5000 base pairs. As used herein, "large insert library" means a gene library containing clones with random large size nucleic acid inserts of approximately 5000 up to several hundred thousand base pairs or greater.

As described with respect to one of the above aspects, the invention provides a process for enzyme activity screening of clones containing selected DNA derived from a microorganism which process includes: screening a library for specified enzyme activity, said library including a plurality of clones, said clones having been prepared by recovering from genomic DNA of a microorganism selected DNA, which DNA is selected by hybridization to at least one DNA sequence which is all or a portion of a DNA sequence encoding an enzyme having the specified activity; and transforming a host with the selected DNA to produce clones which are screened for the specified enzyme activity.

In one embodiment, a DNA library derived from a microorganism is subjected to a selection procedure to select therefrom DNA which hybridizes to one or more probe DNA sequences which is all or a portion of a DNA sequence encoding an enzyme having the specified enzyme activity by: (a) rendering the double-stranded genomic DNA population into a single-stranded DNA population; (b) contacting the single-stranded DNA population of (a) with the DNA probe bound to a ligand under conditions permissive of hybridization so as to produce a double-stranded complex of probe and members of the genomic DNA population which hybridize thereto; (c) contacting the double-stranded complex of (b) with a solid phase specific binding partner for said ligand so as to produce a solid phase complex; (d) separating the solid phase complex from the single-stranded DNA population of (b); (e) releasing from the probe the members of the genomic population which had bound to the solid phase bound probe; (f) forming double-stranded DNA from the members of the genomic population of (e); (g) introducing the double-stranded DNA of (f) into a suitable host to form a library containing a plurality of clones containing the selected DNA; and (h) screening the library for the specified enzyme activity.

In another aspect, the process includes a preselection to recover DNA including signal or secretion sequences. In this manner it is possible to select from the genomic DNA population by hybridization as hereinabove described only DNA which includes a signal or secretion sequence. The following paragraphs describe the protocol for this embodiment of the invention, the nature and function of secretion signal sequences in general and a specific exemplary application of such sequences to an assay or selection process.

A particularly embodiment of this aspect further comprises, after (a) but before (b) above, the steps of: (a*i*) contacting the single-stranded DNA population of (a) with a ligand-bound oligonucleotide probe that is complementary to a secretion signal sequence unique to a given class of proteins under conditions permissive of hybridization to form a double-stranded complex; (a*ii*) contacting the double-stranded complex of (a*i*) with a solid phase specific binding partner for said ligand so as to produce a solid phase complex; (a*iii*) separating the solid phase complex from the single-stranded DNA population of (a); (a*iv*) releasing the members of the genomic population which had bound to said solid phase bound probe; and (a*v*) separating the solid phase bound probe from the members of the genomic population which had bound thereto.

The DNA which has been selected and isolated to include a signal sequence is then subjected to the selection procedure hereinabove described to select and isolate therefrom DNA which binds to one or more probe DNA sequences derived from DNA encoding an enzyme(s) having the specified enzyme activity.

This procedure is described and exemplified in U.S. Serial No. 08/692,002, filed August 2, 1996, incorporated herein by reference.

In vivo biopanning may be performed utilizing a FACS-based and non-optical (*e.g.*, magnetic) based machines. Complex gene libraries are constructed with vectors which contain elements which stabilize transcribed RNA. For example, the inclusion of sequences which result in secondary structures such as hairpins which are designed to flank the transcribed regions of the RNA would serve to enhance their stability, thus

increasing their half life within the cell. The probe molecules used in the biopanning process consist of oligonucleotides labeled with reporter molecules that only fluoresce upon binding of the probe to a target molecule. These probes are introduced into the recombinant cells from the library using one of several transformation methods. The probe molecules bind to the transcribed target mRNA resulting in DNA/RNA heteroduplex molecules. Binding of the probe to a target will yield a fluorescent signal which is detected and sorted by the FACS machine during the screening process.

In some embodiments, the nucleic acid encoding one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, sequences substantially identical thereto, or fragments comprising at least about 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof is assembled in appropriate phase with a leader sequence capable of directing secretion of the translated polypeptide or fragment thereof. Optionally, the nucleic acid encodes a fusion polypeptide in which one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof, is fused to heterologous peptides or polypeptides, such as N-terminal identification peptides which impart desired characteristics, such as increased stability or simplified purification.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is ligated to the desired position in the vector following digestion of the insert and the vector with appropriate restriction endonucleases. Alternatively, blunt ends in both the insert and the vector may be ligated. A variety of cloning techniques are disclosed in Ausubel *et al.* Current Protocols in Molecular Biology, John Wiley 503 Sons, Inc. 1997 and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, 1989, the entire disclosures of which are incorporated herein by reference. Such procedures and others are deemed to be within the scope of those skilled in the art.

The vector may be, for example, in the form of a plasmid, a viral particle, or a phage. Other vectors include chromosomal, nonchromosomal and synthetic DNA sequences, derivatives of SV40; bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. A variety of cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Particular bacterial vectors which may be used include the commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017), pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden), GEM1 (Promega Biotec, Madison, WI, USA) pQE70, pQE60, pQE-9 (Qiagen), pD10, psiX174 pBluescript II KS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene), ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia), pKK232-8 and pCM7. Particular eukaryotic vectors include pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, and pSVL (Pharmacia). However, any other vector may be used as long as it is replicable and viable in the host cell.

The host cell may be any of the host cells familiar to those skilled in the art, including prokaryotic cells, eukaryotic cells, mammalian cells, insect cells, or plant cells. As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as *E. coli*, *Streptomyces*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, fungal cells, such as yeast, insect cells such as *Drosophila S2* and *Spodoptera Sf9*, animal cells such as CHO, COS or Bowes melanoma, and adenoviruses. The selection of an appropriate host is within the abilities of those skilled in the art.

The vector may be introduced into the host cells using any of a variety of techniques, including transformation, transfection, transduction, viral infection, gene guns, or Ti-mediated gene transfer. Particular methods include calcium phosphate

transfection, DEAE-Dextran mediated transfection, lipofection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

Where appropriate, the engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the invention. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter may be induced by appropriate means (*e.g.*, temperature shift or chemical induction) and the cells may be cultured for an additional period to allow them to produce the desired polypeptide or fragment thereof.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract is retained for further purification. Microbial cells employed for expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known to those skilled in the art. The expressed polypeptide or fragment thereof can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the polypeptide. If desired, high performance liquid chromatography (HPLC) can be employed for final purification steps.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts (described by Gluzman, *Cell*, 23:175, 1981), and other cell lines capable of expressing proteins from a compatible vector, such as the C127, 3T3, CHO, HeLa and BHK cell lines.

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Depending upon the host

employed in a recombinant production procedure, the polypeptides produced by host cells containing the vector may be glycosylated or may be non-glycosylated. Polypeptides of the invention may or may not also include an initial methionine amino acid residue. Additional details relating to the recombinant expression of proteins are available to those skilled in the art. For example, Protein Expression : A Practical Approach (Practical Approach Series by S. J. Higgins (Editor), B. D. Hames (Editor) (July 1999) Oxford University Press; ISBN: 0199636249 provides ample guidance to the practioner for the expression of proteins in a wide variety of organisms.

Alternatively, the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof, can be synthetically produced by conventional peptide synthesizers. In other embodiments, fragments or portions of the polypeptides may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides.

As known by those skilled in the art, the nucleic acid sequences of the invention can be optimized for expression in a variety of organisms. In one embodiment, sequences of the invention are optimized for codon usage in an organism of interest, *e.g.*, a fungus such as *S. cerevisiae* or a bacterium such as *E. coli*. Optimization of nucleic acid sequences for the purpose of codon usage is well understood in the art to refer to the selection of a particular codon favored by an organism to encode a particular amino acid. Optimized codon usage tables are known for many organisms. For example, see Transfer RNA in Protein Synthesis by Dolph L. Hatfield, Byeong J. Lee, Robert M. Pirtle (Editor) (July 1992) CRC Press; ISBN: 0849356989. Thus, the invention also includes nucleic acids of the invention adapted for codon usage of an organism.

Optimized expression of nucleic acid sequences of the invention also refers to directed or random mutagenesis of a nucleic acid to effect increased expression of the encoded protein. The mutagenesis of the nucleic acids of the invention can directly or

indirectly provide for an increased yield of expressed protein. By way of non-limiting example, mutagenesis techniques described herein may be utilized to effect mutation of the 5' untranslated region, 3' untranslated region, or coding region of a nucleic acid, the mutation of which can result in increased stability at the RNA or protein level, thereby resulting in an increased yield of protein.

Cell-free translation systems can also be employed to produce one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof, using mRNAs transcribed from a DNA construct comprising a promoter operably linked to a nucleic acid encoding the polypeptide or fragment thereof. In some embodiments, the DNA construct may be linearized prior to conducting an *in vitro* transcription reaction. The transcribed mRNA is then incubated with an appropriate cell-free translation extract, such as a rabbit reticulocyte extract, to produce the desired polypeptide or fragment thereof.

The invention also relates to variants of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14, sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, and 150 consecutive amino acids thereof. The term "variant" includes derivatives or analogs of these polypeptides. In particular, the variants may differ in amino acid sequence from the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, and sequences substantially identical thereto, by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

The variants may be naturally occurring or created *in vitro*. In particular, such variants may be created using genetic engineering techniques such as site directed mutagenesis, random chemical mutagenesis, Exonuclease III deletion procedures, and standard cloning techniques. Alternatively, such variants, fragments, analogs, or derivatives may be created using chemical synthesis or modification procedures.

Other methods of making variants are also familiar to those skilled in the art. These include procedures in which nucleic acid sequences obtained from natural isolates are modified to generate nucleic acids which encode polypeptides having characteristics which enhance their value in industrial or laboratory applications. In such procedures, a large number of variant sequences having one or more nucleotide differences with respect to the sequence obtained from the natural isolate are generated and characterized. Typically, these nucleotide differences result in amino acid changes with respect to the polypeptides encoded by the nucleic acids from the natural isolates.

For example, variants may be created using error prone PCR. In error prone PCR, PCR is performed under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. Error prone PCR is described in Leung, D.W., *et al.*, *Technique*, 1:11-15, 1989) and Caldwell, R. C. and Joyce G.F., PCR *Methods Applic.*, 2:28-33, 1992, the disclosure of which is incorporated herein by reference in its entirety. Briefly, in such procedures, nucleic acids to be mutagenized are mixed with PCR primers, reaction buffer, $MgCl_2$, $MnCl_2$, Taq polymerase and an appropriate concentration of dNTPs for achieving a high rate of point mutation along the entire length of the PCR product. For example, the reaction may be performed using 20 fmoles of nucleic acid to be mutagenized, 30pmole of each PCR primer, a reaction buffer comprising 50 mM KCl, 10 mM Tris HCl (pH 8.3) and 0.01% gelatin, 7mM $MgCl_2$, 0.5 mM $MnCl_2$, 5 units of Taq polymerase, 0.2 mM dGTP, 0.2 mM dATP, 1 mM dCTP, and 1 mM dTTP. PCR may be performed for 30 cycles of 94° C for 1 min, 45° C for 1 min, and 72° C for 1 min. However, it will be appreciated that these parameters may be varied as appropriate. The mutagenized nucleic acids are cloned into an appropriate vector and the activities of the polypeptides encoded by the mutagenized nucleic acids is evaluated.

Variants may also be created using oligonucleotide directed mutagenesis to generate site-specific mutations in any cloned DNA of interest. Oligonucleotide mutagenesis is described in Reidhaar-Olson, J.F. and Sauer, R.T., *et al.*, *Science*, 241:53-57, 1988, the disclosure of which is incorporated herein by reference in its entirety. Briefly, in such procedures a plurality of double stranded oligonucleotides

bearing one or more mutations to be introduced into the cloned DNA are synthesized and inserted into the cloned DNA to be mutagenized. Clones containing the mutagenized DNA are recovered and the activities of the polypeptides they encode are assessed.

Another method for generating variants is assembly PCR. Assembly PCR involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction. Assembly PCR is described in pending U.S. Patent Application Serial No. 08/677,112 filed July 9, 1996, entitled, Method of "DNA Shuffling with Polynucleotides Produced by Blocking or interrupting a Synthesis or Amplification Process," the disclosure of which is incorporated herein by reference in its entirety.

Still another method of generating variants is sexual PCR mutagenesis. In sexual PCR mutagenesis, forced homologous recombination occurs between DNA molecules of different but highly related DNA sequence *in vitro*, as a result of random fragmentation of the DNA molecule based on sequence homology, followed by fixation of the crossover by primer extension in a PCR reaction. Sexual PCR mutagenesis is described in Stemmer, W.P., *PNAS*, USA, 91:10747-10751, 1994, the disclosure of which is incorporated herein by reference. Briefly, in such procedures a plurality of nucleic acids to be recombined are digested with DNAse to generate fragments having an average size of 50-200 nucleotides. Fragments of the desired average size are purified and resuspended in a PCR mixture. PCR is conducted under conditions which facilitate recombination between the nucleic acid fragments. For example, PCR may be performed by resuspending the purified fragments at a concentration of 10-30ng/μl in a solution of 0.2 mM of each dNTP, 2.2 mM MgCl2, 50 mM KCL, 10 mM Tris HCl, pH 9.0, and 0.1% Triton X-100. 2.5 units of Taq polymerase per 100μl of reaction mixture is added and PCR is performed using the following regime: 94° C for 60 seconds, 94° C for 30 seconds, 50-55° C for 30 seconds, 72° C for 30 seconds (30-45 times) and 72° C for 5 minutes. However, it will be appreciated that these parameters may be varied as appropriate. In some embodiments, oligonucleotides may be included in the PCR reactions. In other embodiments, the Klenow fragment of DNA polymerase I may be

used in a first set of PCR reactions and Taq polymerase may be used in a subsequent set of PCR reactions. Recombinant sequences are isolated and the activities of the polypeptides they encode are assessed.

Variants may also be created by *in vivo* mutagenesis. In some embodiments, random mutations in a sequence of interest are generated by propagating the sequence of interest in a bacterial strain, such as an E. coli strain, which carries mutations in one or more of the DNA repair pathways. Such "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in one of these strains will eventually generate random mutations within the DNA. Mutator strains suitable for use for *in vivo* mutagenesis are described in PCT Publication No. WO 91/16427, published October 31, 1991, entitled "Methods for Phenotype Creation from Multiple Gene Populations" the disclosure of which is incorporated herein by reference in its entirety.

Variants may also be generated using cassette mutagenesis. In cassette mutagenesis a small region of a double stranded DNA molecule is replaced with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

Recursive ensemble mutagenesis may also be used to generate variants. Recursive ensemble mutagenesis is an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. Recursive ensemble mutagenesis is described in Arkin, A.P. and Youvan, D.C., *PNAS*, USA, 89:7811-7815, 1992, the disclosure of which is incorporated herein by reference in its entirety.

In some embodiments, variants are created using exponential ensemble mutagenesis. Exponential ensemble mutagenesis is a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered

position, amino acids which lead to functional proteins. Exponential ensemble mutagenesis is described in Delegrave, S. and Youvan, D.C., *Biotechnol. Res.*, 11:1548-1552, 1993, the disclosure of which incorporated herein by reference in its entirety. Random and site-directed mutagenesis are described in Arnold, F.H., *Current Opinion in Biotechnology*, 4:450-455, 1993, the disclosure of which is incorporated herein by reference in its entirety.

In some embodiments, the variants are created using shuffling procedures wherein portions of a plurality of nucleic acids which encode distinct polypeptides are fused together to create chimeric nucleic acid sequences which encode chimeric polypeptides as described in pending U.S. Patent Application Serial No. 08/677,112 filed July 9, 1996, entitled, "Method of DNA Shuffling with Polynucleotides Produced by Blocking or interrupting a Synthesis or Amplification Process", and pending U.S. Patent Application Serial No. 08/651,568 filed May 22, 1996, entitled, "Combinatorial Enzyme Development."

The variants of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 may be variants in which one or more of the amino acid residues of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code.

Conservative substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the following replacements: replacements of an aliphatic amino acid such as Ala, Val, Leu and Ile with another aliphatic amino acid; replacement of a Ser with a Thr or vice versa; replacement of an acidic residue such as Asp and Glu with another acidic residue; replacement of a residue bearing an amide group, such as Asn and Gln, with another residue bearing an amide group; exchange of a basic residue such as Lys and Arg with another basic residue; and replacement of an aromatic residue such as Phe, Tyr with another aromatic residue.

Other variants are those in which one or more of the amino acid residues of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 includes a substituent group.

Still other variants are those in which the polypeptide is associated with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol).

Additional variants are those in which additional amino acids are fused to the polypeptide, such as a leader sequence, a secretory sequence, a proprotein sequence or a sequence which facilitates purification, enrichment, or stabilization of the polypeptide.

In some embodiments, the fragments, derivatives and analogs retain the same biological function or activity as the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, and sequences substantially identical thereto. In other embodiments, the fragment, derivative, or analog includes a proprotein, such that the fragment, derivative, or analog can be activated by cleavage of the proprotein portion to produce an active polypeptide.

Another aspect of the invention is polypeptides or fragments thereof which have at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or more than about 95% homology to one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, sequences substantially identical thereto, or a fragment comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof. Homology may be determined using any of the programs described above which aligns the polypeptides or fragments being compared and determines the extent of amino acid identity or similarity between them. It will be appreciated that amino acid "homology" includes conservative amino acid substitutions such as those described above.

The polypeptides or fragments having homology to one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID

NO:12 or SEQ ID NO:14, sequences substantially identical thereto, or a fragment comprising at least about 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof, may be obtained by isolating the nucleic acids encoding them using the techniques described above.

Alternatively, the homologous polypeptides or fragments may be obtained through biochemical enrichment or purification procedures. The sequence of potentially homologous polypeptides or fragments may be determined by proteolytic digestion, gel electrophoresis and/or microsequencing. The sequence of the prospective homologous polypeptide or fragment can be compared to one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, sequences substantially identical thereto, or a fragment comprising at least about 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof using any of the programs described herein.

Another aspect of the invention is an assay for identifying fragments or variants of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, or sequences substantially identical thereto, which retain the enzymatic function of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 and sequences substantially identical thereto. For example the fragments or variants of the polypeptides, may be used to catalyze biochemical reactions, which indicate that said fragment or variant retains the enzymatic activity of the polypeptides in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14.

The assay for determining if fragments of variants retain the enzymatic activity of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 and sequences substantially identical thereto includes the steps of; contacting the polypeptide fragment or variant with a substrate molecule under conditions which allow the polypeptide fragment or variant to function, and detecting either a decrease in the level of substrate or an increase in the

level of the specific reaction product of the reaction between the polypeptide and substrate.

The polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof, may be used in a variety of applications. For example, the polypeptides or fragments thereof may be used to catalyze biochemical reactions. In accordance with one aspect of the invention, there is provided a process for utilizing a polypeptide having SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 and sequences substantially identical thereto, or polynucleotides encoding such polypeptides for hydrolyzing haloalkanes. In such procedures, a substance containing a haloalkane compound is contacted with one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, or SEQ ID NO:14, sequences substantially identical thereto, under conditions which facilitate the hydrolysis of the compound.

The polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof, may also be used to generate antibodies which bind specifically to the enzyme polypeptides or fragments. The resulting antibodies may be used in immunoaffinity chromatography procedures to isolate or purify the polypeptide or to determine whether the polypeptide is present in a biological sample. In such procedures, a protein preparation, such as an extract, or a biological sample is contacted with an antibody capable of specifically binding to one of a polypeptide of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 sequences substantially identical thereto, or fragments of the foregoing sequences.

In immunoaffinity procedures, the antibody is attached to a solid support, such as a bead or other column matrix. The protein preparation is placed in contact with the

antibody under conditions in which the antibody specifically binds to one of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, sequences substantially identical thereto, or fragment thereof. After a wash to remove non-specifically bound proteins, the specifically bound polypeptides are eluted.

The ability of proteins in a biological sample to bind to the antibody may be determined using any of a variety of procedures familiar to those skilled in the art. For example, binding may be determined by labeling the antibody with a detectable label such as a fluorescent agent, an enzymatic label, or a radioisotope. Alternatively, binding of the antibody to the sample may be detected using a secondary antibody having such a detectable label thereon. Particular assays include ELISA assays, sandwich assays, radioimmunoassays, and Western Blots.

Polyclonal antibodies generated against the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, and sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof, can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, for example, a non-human. The antibody so obtained then binds the polypeptide itself. In this manner, even a sequence encoding only a fragment of the polypeptide can be used to generate antibodies which may bind to the whole native polypeptide. Such antibodies can then be used to isolate the polypeptide from cells expressing that polypeptide.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, *Nature*, 256:495-497, 1975, the disclosure of which is incorporated herein by reference), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunol. Today* 4:72, 1983, the disclosure of which is incorporated herein by reference), and the EBV-hybridoma technique (Cole, *et al.*, 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96, the disclosure of which is incorporated herein by reference).

Techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778, the disclosure of which is incorporated herein by reference) can be adapted to produce single chain antibodies to the polypeptides of, for example, SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 and fragments thereof. Alternatively, transgenic mice may be used to express humanized antibodies to these polypeptides or fragments.

Antibodies generated against a polypeptide of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, sequences substantially identical thereto, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof, may be used in screening for similar polypeptides from other organisms and samples. In such techniques, polypeptides from the organism are contacted with the antibody and those polypeptides which specifically bind the antibody are detected. Any of the procedures described above may be used to detect antibody binding. One such screening assay is described in "Methods for Measuring Cellulase Activities", *Methods in Enzymology*, Vol 160, pp. 87-116, which is hereby incorporated by reference in its entirety.

As used herein the term "nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13" encompasses a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, a sequence substantially identical to one of the foregoing sequences, fragments of any one or more of the foregoing sequences, nucleotide sequences homologous to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, or homologous to fragments of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences complementary to all of the preceding sequences. The fragments include portions of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13 comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive nucleotides of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences

substantially identical thereto. Homologous sequences and fragments of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto, refer to a sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75% or 70% homology to these sequences. Homology may be determined using any of the computer programs and parameters described herein, including FASTA version 3.0t78 with the default parameters. Homologous sequences also include RNA sequences in which uridines replace the thymines in the nucleic acid sequences as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13. The homologous sequences may be obtained using any of the procedures described herein or may result from the correction of a sequencing error. It will be appreciated that the nucleic acid sequences of the invention can be represented in the traditional single character format (See the inside back cover of Stryer, Lubert. *Biochemistry*, 3rd edition. W. H Freeman and Co., New York.) or in any other format which records the identity of the nucleotides in a sequence.

As used herein the term "a polypeptide sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14" encompasses s polypeptide sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, sequences substantially identical thereto, which are encoded by a sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, polypeptide sequences homologous to the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, and sequences substantially identical thereto, or fragments of any of the preceding sequences. Homologous polypeptide sequences refer to a polypeptide sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75% or 70% homology to one of the polypeptide sequences of the invention. Homology may be determined using any of the computer programs and parameters described herein, including FASTA version 3.0t78 with the default parameters or with any modified parameters. The homologous sequences may be obtained using any of the procedures described herein or may result from the correction of a sequencing error. The polypeptide fragments comprise at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or

150 consecutive amino acids of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, and sequences substantially identical thereto. It will be appreciated that the polypeptides of the invention can be represented in the traditional single character format or three letter format (See the inside back cover of Starrier, Lubert. *Biochemistry*, 3rd edition. W. H Freeman and Co., New York.) or in any other format which relates the identity of the polypeptides in a sequence.

It will be appreciated by those skilled in the art that a nucleic acid sequence and a polypeptide sequence of the invention can be stored, recorded, and manipulated on any medium which can be read and accessed by a computer. As used herein, the words "recorded" and "stored" refer to a process for storing information on a computer medium. A skilled artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate manufactures comprising one or more of the nucleic acid sequences as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto, one or more of the polypeptide sequences as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14, and sequences substantially identical thereto. Another aspect of the invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, or 20 nucleic acid sequences as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, and sequences substantially identical thereto.

Another aspect of the invention is a computer readable medium having recorded thereon one or more of the nucleic acid sequences as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto. Another aspect of the invention is a computer readable medium having recorded thereon one or more of the polypeptide sequences as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, and sequences substantially identical thereto. Another aspect of the invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, or 20 of the sequences as set forth above.

Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disk, a floppy disk, a magnetic tape, CD-ROM, Digital Versatile Disk (DVD), Random Access Memory (RAM), or Read Only Memory (ROM) as well as other types of other media known to those skilled in the art.

Embodiments of the invention include systems (*e.g.*, internet based systems), particularly computer systems which store and manipulate the sequence information described herein. One example of a computer system 100 is illustrated in block diagram form in Figure 1. As used herein, "a computer system" refers to the hardware components, software components, and data storage components used to analyze a nucleotide sequence of a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto, or a polypeptide sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14. The computer system 100 typically includes a processor for processing, accessing and manipulating the sequence data. The processor 105 can be any well-known type of central processing unit, such as, for example, the Pentium III from Intel Corporation, or similar processor from Sun, Motorola, Compaq, AMD or International Business Machines.

Typically the computer system 100 is a general purpose system that comprises the processor 105 and one or more internal data storage components 110 for storing data, and one or more data retrieving devices for retrieving the data stored on the data storage components. A skilled artisan can readily appreciate that any one of the currently available computer systems are suitable.

In one particular embodiment, the computer system 100 includes a processor 105 connected to a bus which is connected to a main memory 115 (preferably implemented as RAM) and one or more internal data storage devices 110, such as a hard drive and/or other computer readable media having data recorded thereon. In some embodiments, the

computer system 100 further includes one or more data retrieving device 118 for reading the data stored on the internal data storage devices 110.

The data retrieving device 118 may represent, for example, a floppy disk drive, a compact disk drive, a magnetic tape drive, or a modem capable of connection to a remote data storage system (*e.g.*, via the internet) etc. In some embodiments, the internal data storage device 110 is a removable computer readable medium such as a floppy disk, a compact disk, a magnetic tape, etc. containing control logic and/or data recorded thereon. The computer system 100 may advantageously include or be programmed by appropriate software for reading the control logic and/or the data from the data storage component once inserted in the data retrieving device.

The computer system 100 includes a display 120 which is used to display output to a computer user. It should also be noted that the computer system 100 can be linked to other computer systems 125a-c in a network or wide area network to provide centralized access to the computer system 100.

Software for accessing and processing the nucleotide sequences of a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto, or a polypeptide sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 and sequences substantially identical thereto, (such as search tools, compare tools, and modeling tools etc.) may reside in main memory 115 during execution.

In some embodiments, the computer system 100 may further comprise a sequence comparison algorithm for comparing a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto, or a polypeptide sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, and sequences substantially identical thereto, stored on a computer readable medium to a reference nucleotide or polypeptide sequence(s) stored on a computer readable medium. A "sequence comparison algorithm"

refers to one or more programs which are implemented (locally or remotely) on the computer system 100 to compare a nucleotide sequence with other nucleotide sequences and/or compounds stored within a data storage means. For example, the sequence comparison algorithm may compare the nucleotide sequences of a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto, or a polypeptide sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, and sequences substantially identical thereto, stored on a computer readable medium to reference sequences stored on a computer readable medium to identify homologies or structural motifs. Various sequence comparison programs identified elsewhere in this patent specification are particularly contemplated for use in this aspect of the invention. Protein and/or nucleic acid sequence homologies may be evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, *Proc. Natl. Acad. Sci.* USA 85(8):2444-2448, 1988; Altschul *et al.*, *J. Mol. Biol.* 215(3):403-410, 1990; Thompson *et al.*, *Nucleic Acids Res.* 22(2):4673-4680, 1994; Higgins *et al.*, *Methods Enzymol.* 266:383-402, 1996; Altschul *et al.*, *J. Mol. Biol.* 215(3):403-410, 1990; Altschul *et al.*, *Nature Genetics* 3:266-272, 1993).

Homology or identity is often measured using sequence analysis software (*e.g.*, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). Such software matches similar sequences by assigning degrees of homology to various deletions, substitutions and other modifications. The terms "homology" and "identity" in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same when compared and aligned for maximum correspondence over a comparison window or designated region as measured using any number of sequence comparison algorithms or by manual alignment and visual inspection.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequence for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith and Waterman, *Adv. Appl. Math.* 2:482, 1981, by the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol* 48:443, 1970, by the search for similarity method of person and Lipman, *Proc. Nat'l. Acad. Sci.* USA 85:2444, 1988, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection. Other algorithms for determining homology or identity include, for example, in addition to a BLAST program (Basic Local Alignment Search Tool at the National Center for Biological Information), ALIGN, AMAS (Analysis of Multiply Aligned Sequences), AMPS (Protein Multiple Sequence Alignment), ASSET (Aligned Segment Statistical Evaluation Tool), BANDS, BESTSCOR, BIOSCAN (Biological Sequence Comparative Analysis Node), BLIMPS (BLocks IMProved Searcher), FASTA, Intervals and Points, BMB, CLUSTAL V, CLUSTAL W, CONSENSUS, LCONSENSUS, WCONSENSUS, Smith-Waterman algorithm, DARWIN, Las Vegas algorithm, FNAT (Forced Nucleotide Alignment Tool), Framealign, Framesearch, DYNAMIC, FILTER, FSAP (Fristensky Sequence Analysis Package), GAP (Global Alignment Program), GENAL, GIBBS, GenQuest, ISSC (Sensitive Sequence Comparison), LALIGN (Local Sequence Alignment), LCP (Local Content Program), MACAW (Multiple Alignment Construction and Analysis Workbench), MAP (Multiple Alignment Program), MBLKP, MBLKN, PIMA (Pattern-

Induced Multi-sequence Alignment), SAGA (Sequence Alignment by Genetic Algorithm) and WHAT-IF. Such alignment programs can also be used to screen genome databases to identify polynucleotide sequences having substantially identical sequences. A number of genome databases are available, for example, a substantial portion of the human genome is available as part of the Human Genome Sequencing Project (J. Roach, http://weber.u.Washington.edu/~roach/human_ genome_ progress 2.html) (Gibbs, 1995). At least twenty-one other genomes have already been sequenced, including, for example, *M. genitalium* (Fraser *et al.*, 1995), M. jannaschii (Bult *et al.*, 1996), *H. influenzae* (Fleischmann *et al.*, 1995), *E. coli* (Blattner *et al.*, 1997), and yeast (*S. cerevisiae*) (Mewes *et al.*, 1997), and *D. melanogaster* (Adams *et al.*, 2000). Significant progress has also been made in sequencing the genomes of model organism, such as mouse, *C. elegans*, and *Arabadopsis sp.* Several databases containing genomic information annotated with some functional information are maintained by different organization, and are accessible via the internet, for example, http://wwwtigr.org/tdb; http://www.genetics.wisc.edu; http://genome-www.stanford.edu/~ball; http://hiv-web.lanl.gov; http://www.ncbi.nlm.nih.gov; http://www.ebi.ac.uk; http://Pasteur.fr/other/biology; and http:// www.genome.wi.mit.edu.

One example of a useful algorithm is BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.*, *Nuc. Acids Res.* 25:3389-3402, 1977, and Altschul *et al.*, *J. Mol. Biol.* 215:403-410, 1990, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off

by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectations (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff, *Proc. Natl. Acad. Sci.* USA 89:10915, 1989) alignments (B) of 50, expectation (E) of 10, M=5, N= -4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, *e.g.*, Karlin and Altschul, *Proc. Natl. Acad. Sci.* USA 90:5873, 1993). One measure of similarity provided by BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a references sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

In one embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") In particular, five specific BLAST programs are used to perform the following task:

(1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;

(2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;

(3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;

(4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

(5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (*i.e.*, aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet *et al., Science* 256:1443-1445, 1992; Henikoff and Henikoff, *Proteins* 17:49-61, 1993). Less preferably, the PAM or PAM250 matrices may also be used (see, *e.g.*, Schwartz and Dayhoff, eds., 1978, *Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure*, Washington: National Biomedical Research Foundation). BLAST programs are accessible through the U.S. National Library of Medicine, *e.g.*, at www.ncbi.nlm.nih.gov.

The parameters used with the above algorithms may be adapted depending on the sequence length and degree of homology studied. In some embodiments, the parameters may be the default parameters used by the algorithms in the absence of instructions from the user.

Figure 2 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database. The database of sequences can be a private database stored within the computer system 100, or a public database such as GENBANK that is available through the Internet.

The process 200 begins at a start state 201 and then moves to a state 202 wherein the new sequence to be compared is stored to a memory in a computer system 100. As discussed above, the memory could be any type of memory, including RAM or an internal storage device.

The process 200 then moves to a state 204 wherein a database of sequences is opened for analysis and comparison. The process 200 then moves to a state 206 wherein the first sequence stored in the database is read into a memory on the computer. A comparison is then performed at a state 210 to determine if the first sequence is the same as the second sequence. It is important to note that this step is not limited to performing an exact comparison between the new sequence and the first sequence in the database. Well-known methods are known to those of skill in the art for comparing two nucleotide or protein sequences, even if they are not identical. For example, gaps can be introduced into one sequence in order to raise the homology level between the two tested sequences. The parameters that control whether gaps or other features are introduced into a sequence during comparison are normally entered by the user of the computer system.

Once a comparison of the two sequences has been performed at the state 210, a determination is made at a decision state 210 whether the two sequences are the same. Of course, the term "same" is not limited to sequences that are absolutely identical. Sequences that are within the homology parameters entered by the user will be marked as "same" in the process 200.

If a determination is made that the two sequences are the same, the process 200 moves to a state 214 wherein the name of the sequence from the database is displayed to the user. This state notifies the user that the sequence with the displayed name fulfills the homology constraints that were entered. Once the name of the stored sequence is displayed to the user, the process 200 moves to a decision state 218 wherein a determination is made whether more sequences exist in the database. If no more sequences exist in the database, then the process 200 terminates at an end state 220. However, if more sequences do exist in the database, then the process 200 moves to a state 224 wherein a pointer is moved to the next sequence in the database so that it can be compared to the new sequence. In this manner, the new sequence is aligned and compared with every sequence in the database.

It should be noted that if a determination had been made at the decision state 212 that the sequences were not homologous, then the process 200 would move immediately

to the decision state 218 in order to determine if any other sequences were available in the database for comparison.

Accordingly, one aspect of the invention is a computer system comprising a processor, a data storage device having stored thereon a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto, or a polypeptide sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 and sequences substantially identical thereto, a data storage device having retrievably stored thereon reference nucleotide sequences or polypeptide sequences to be compared to a nucleic acid sequence or a polypeptide sequence of the invention, and a sequence comparer for conducting the comparison. The sequence comparer may indicate a homology level between the sequences compared or identify structural motifs in the above described nucleic acid code of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, and sequences substantially identical thereto, or a polypeptide sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 and sequences substantially identical thereto, or it may identify structural motifs in sequences which are compared to these nucleic acid codes and polypeptide codes. In some embodiments, the data storage device may have stored thereon the sequences of at least 2, 5, 10, 15, 20, 25, 30 or 40 or more of the nucleic acid sequences as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, and sequences substantially identical thereto, or the polypeptide sequences as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14, and sequences substantially identical thereto.

Another aspect of the invention is a method for determining the level of homology between a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto, or a polypeptide sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID

NO:14 and sequences substantially identical thereto, and a reference nucleotide sequence. The method including reading the nucleic acid code or the polypeptide code and the reference nucleotide or polypeptide sequence through the use of a computer program which determines homology levels and determining homology between the nucleic acid code or polypeptide code and the reference nucleotide or polypeptide sequence with the computer program. The computer program may be any of a number of computer programs for determining homology levels, including those specifically enumerated herein, (*e.g.*, BLAST2N with the default parameters or with any modified parameters). The method may be implemented using the computer systems described above. The method may also be performed by reading at least 2, 5, 10, 15, 20, 25, 30 or 40 or more of the above described nucleic acid sequences as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, or the polypeptide sequences as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14 through use of the computer program and determining homology between the nucleic acid codes or polypeptide codes and reference nucleotide sequences or polypeptide sequences.

Figure 3 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous. The process 250 begins at a start state 252 and then moves to a state 254 wherein a first sequence to be compared is stored to a memory. The second sequence to be compared is then stored to a memory at a state 256. The process 250 then moves to a state 260 wherein the first character in the first sequence is read and then to a state 262 wherein the first character of the second sequence is read. It should be understood that if the sequence is a nucleotide sequence, then the character would normally be either A, T, C, G or U. If the sequence is a protein sequence, then it is preferably in the single letter amino acid code so that the first and sequence sequences can be easily compared.

A determination is then made at a decision state 264 whether the two characters are the same. If they are the same, then the process 250 moves to a state 268 wherein the next characters in the first and second sequences are read. A determination is then made whether the next characters are the same. If they are, then the process 250

continues this loop until two characters are not the same. If a determination is made that the next two characters are not the same, the process 250 moves to a decision state 274 to determine whether there are any more characters either sequence to read.

If there are not any more characters to read, then the process 250 moves to a state 276 wherein the level of homology between the first and second sequences is displayed to the user. The level of homology is determined by calculating the proportion of characters between the sequences that were the same out of the total number of sequences in the first sequence. Thus, if every character in a first 100 nucleotide sequence aligned with a every character in a second sequence, the homology level would be 100%.

Alternatively, the computer program may be a computer program which compares the nucleotide sequences of a nucleic acid sequence as set forth in the invention, to one or more reference nucleotide sequences in order to determine whether the nucleic acid code of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, and sequences substantially identical thereto, differs from a reference nucleic acid sequence at one or more positions. Optionally such a program records the length and identity of inserted, deleted or substituted nucleotides with respect to the sequence of either the reference polynucleotide or a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto. In one embodiment, the computer program may be a program which determines whether a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto, contains a single nucleotide polymorphism (SNP) with respect to a reference nucleotide sequence.

Accordingly, another aspect of the invention is a method for determining whether a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto, differs at one or more nucleotides from a reference nucleotide sequence comprising the steps of reading the nucleic acid code and the

reference nucleotide sequence through use of a computer program which identifies differences between nucleic acid sequences and identifying differences between the nucleic acid code and the reference nucleotide sequence with the computer program. In some embodiments, the computer program is a program which identifies single nucleotide polymorphisms. The method may be implemented by the computer systems described above and the method illustrated in Figure 3. The method may also be performed by reading at least 2, 5, 10, 15, 20, 25, 30, or 40 or more of the nucleic acid sequences as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13, and sequences substantially identical thereto, and the reference nucleotide sequences through the use of the computer program and identifying differences between the nucleic acid codes and the reference nucleotide sequences with the computer program.

In other embodiments the computer based system may further comprise an identifier for identifying features within a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, or a polypeptide sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, and sequences substantially identical thereto.

An "identifier" refers to one or more programs which identifies certain features within a nucleic acid sequence or a polypeptide sequence of the invention. In one embodiment, the identifier may comprise a program which identifies an open reading frame in a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto.

In another aspect, the invention provides a method to identity a phytate sequence comprising analyzing an amino acid sequence for the occurrence of a first region consisting of RHGVRXaaPT and a second region consisting of WPXaaWPV, wherein the first and second region are separated by 13 amino acids. In various embodiments thereof, the first and the second region are separated by 10, 11, 12, 14, 15, and 16 amino acids.

Figure 5 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence. The process 300 begins at a start state 302 and then moves to a state 304 wherein a first sequence that is to be checked for features is stored to a memory 115 in the computer system 100. The process 300 then moves to a state 306 wherein a database of sequence features is opened. Such a database would include a list of each feature's attributes along with the name of the feature. For example, a feature name could be "Initiation Codon" and the attribute would be "ATG". Another example would be the feature name "TAATAA Box" and the feature attribute would be "TAATAA". An example of such a database is produced by the University of Wisconsin Genetics Computer Group (www.gcg.com). Alternatively, the features may be structural polypeptide motifs such as alpha helices, beta sheets, or functional polypeptide motifs such as enzymatic active sites, helix-turn-helix motifs or other motifs known to those skilled in the art.

Once the database of features is opened at the state 306, the process 300 moves to a state 308 wherein the first feature is read from the database. A comparison of the attribute of the first feature with the first sequence is then made at a state 310. A determination is then made at a decision state 316 whether the attribute of the feature was found in the first sequence. If the attribute was found, then the process 300 moves to a state 318 wherein the name of the found feature is displayed to the user.

The process 300 then moves to a decision state 320 wherein a determination is made whether move features exist in the database. If no more features do exist, then the process 300 terminates at an end state 324. However, if more features do exist in the database, then the process 300 reads the next sequence feature at a state 326 and loops back to the state 310 wherein the attribute of the next feature is compared against the first sequence.

It should be noted, that if the feature attribute is not found in the first sequence at the decision state 316, the process 300 moves directly to the decision state 320 in order to determine if any more features exist in the database.

Accordingly, another aspect of the invention is a method of identifying a feature within a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto, or a polypeptide sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 and sequences substantially identical thereto, comprising reading a nucleic acid sequence or a polypeptide sequence through the use of a computer program which identifies features therein and identifying features within the nucleic acid sequence or polypeptide sequence with the computer program. In one embodiment, computer program comprises a computer program which identifies open reading frames. The method may be performed by reading a single sequence or at least 2, 5, 10, 15, 20, 25, 30, or 40 of the nucleic acid sequences as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto, or the polypeptide sequences as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 and sequences substantially identical thereto, through the use of the computer program and identifying features within the nucleic acid codes or polypeptide codes with the computer program.

In addition, a nucleic acid sequence or a polypeptide sequence of the invention may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto, or a polypeptide sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 and sequences substantially identical thereto, may be stored as text in a word processing file, such as MicrosoftWORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE. In addition, many computer programs and databases may be used as sequence comparison algorithms, identifiers, or sources of reference nucleotide sequences or polypeptide sequences to be compared to a nucleic acid sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, and sequences substantially identical thereto,

or a polypeptide sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 and sequences substantially identical thereto. The following list is intended not to limit the invention but to provide guidance to programs and databases which are useful with the nucleic acid sequences or the polypeptide sequences of the invention.

The programs and databases which may be used include, but are not limited to: MacPattern (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul et al, J. Mol. Biol. 215: 403, 1990), FASTA (Pearson and Lipman, Proc. Natl. Acad. Sci. USA, 85: 2444, 1988), FASTDB (Brutlag *et al.* Comp. App. Biosci. 6:237-245, 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius^2.DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMm (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), the MDL Available Chemicals Directory database, the MDL Drug Data Report data base, the Comprehensive Medicinal Chemistry database, Derwents's World Drug Index database, the BioByteMasterFile database, the Genbank database, and the Genseqn database. Many other programs and data bases would be apparent to one of skill in the art given the present disclosure.

Motifs which may be detected using the above programs include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

The isolated polynucleotide sequences, polypeptide sequence, variants and mutants thereof can be measured for retention of biological activity characteristic to the enzyme of the present invention, for example, in an assay for detecting enzymatic phytase activity (Food Chemicals Codex, 4th Ed.). Such enzymes include truncated forms of phytase, and variants such as deletion and insertion variants of the polypeptide sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14.

An *in vitro* example of such an assay is the following assay for the detection of phytase activity: Phytase activity can be measured by incubating 150µl of the enzyme preparation with 600 µl of 2 mM sodium phytate in 100 mM Tris HCl buffer, pH 7.5, supplemented with 1 mM $CaCl_2$ for 30 minutes at 37°C. After incubation the reaction is stopped by adding 750 µl of 5% trichloroacetic acid. Phosphate released was measured against phosphate standard spectrophotometrically at 700nm after adding 1500 µl of the color reagent (4 volumes of 1.5% ammonium molybdate in 5.5% sulfuric acid and 1 volume of 2.7% ferrous sulfate; Shimizu, 1992). One unit of enzyme activity is defined as the amount of enzyme required to liberate one µmol Pi per min under assay conditions. Specific activity can be expressed in units of enzyme activity per mg of protein. The enzyme of the present invention has enzymatic activity with respect to the hydrolysis of phytate to inositol and free phosphate.

In one embodiment, the instant invention provides a method (and products thereof) of producing stabilized aqueous liquid formulations having phytase activity that exhibit increased resistance to heat inactivation of the enzyme activity and which retain their phytase activity during prolonged periods of storage. The liquid formulations are stabilized by means of the addition of urea and/or a polyol such as sorbitol and glycerol as stabilizing agent. Also provided are feed preparations for monogastric animals and methods for the production thereof that result from the use of such stabilized aqueous liquid formulations. Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes EP 0626010 (W0

9316175 A1) (Barendse *et al.*), although references in the publicly available literature do not teach the inventive molecules of the instant application.

In one embodiment, the instant invention provides a method of hydrolyzing phytate comprised of contacting the phytate with one or more of the novel phytase molecules disclosed herein (*e.g.*, SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, or SEQ ID NO:14). Accordingly, the invention provides a method for catalyzing the hydrolysis of phytate to inositol and free phosphate with release of minerals from the phytic acid complex. The method includes contacting a phytate substrate with a degrading effective amount of an enzyme of the invention, such as the enzyme shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, or SEQ ID NO:14. The term "degrading effective" amount refers to the amount of enzyme which is required to degrade at least 50% of the phytate, as compared to phytate not contacted with the enzyme. Preferably, at least 80% of the phytate is degraded.

In another embodiment, the invention provides a method for hydrolyzing phospho-mono-ester bonds in phytate. The method includes administering an effective amount of phytase molecules of the invention (*e.g.*, SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, or SEQ ID NO:14), to yield inositol and free phosphate. An "effective" amount refers to the amount of enzyme which is required to hydrolyze at least 50% of the phospho-mono-ester bonds, as compared to phytate not contacted with the enzyme. Preferably, at least 80% of the bonds are hydrolyzed.

In a particular aspect, when desired, the phytase molecules may be used in combination with other reagents, such as other catalysts; in order to effect chemical changes (*e.g.* hydrolysis) in the phytate molecules and/or in other molecules of the substrate source(s). According to this aspect, preferably the phytase molecules and the additional reagent(s) will not inhibit each other, more preferably the phytase molecules and the additional reagent(s) will have an overall additive effect, and more preferably still the phytase molecules and the additional reagent(s) will have an overall synergistic effect.

Relevant sources of the substrate phytate molecules include foodstuffs, potential foodstuffs, byproducts of foodstuffs (both *in vitro* byproducts and *in vivo* byproducts, *e.g.* ex vivo reaction products and animal excremental products), precursors of foodstuffs, and any other material source of phytate.

In a non-limiting apsect, the recombinant phytase can be consumed by organisms and retains activity upon consumption. In another exemplification, transgenic approches can be used to achieve expression of the recombinant phytase – preferably in a controlled fashion (methods are available for controlling expression of transgenic molecules in time-specific and tissue specific manners).

In a particular exemplification, the phytase activity in the source material (*e.g.* a transgenic plant source or a recombinant prokaryotic host) may be increased upon consumption; this increase in activity may occur, for example, upon conversion of a precursor phytase molecule in pro-form to a significantly more active enzyme in a more mature form, where said conversion may result, for example, from the injestion and digestion of the phytase source. Hydrolysis of the phytate substrate may occur at any time upon the contacting of the phytase with the phytate; for example, this may occur before injestion or after injestion or both before and after injestion of either the substrate or the enzyme or both. It is additionally appreciated that the phytate substrate may be contacted with – in addition to the phytase – one or more additional reagents, such as another enzyme, which may be also be applied either directly or after purification from its source material.

It is appreciated that the phytase source material(s) can be contacted directly with the phytate source material(s); *e.g.* upon *in vitro* or *in vivo* grinding or chewing of either or both the phytase source(s) and the phytate source(s). Alternatively the phytase enzyme may be purified away from source material(s), or the phytate substrate may be purified away from source material(s), or both the phytase enzyme and the phytate substrate may be purified away from source material(s) prior to the contacing of the phytase enzyme with the phytate substrate. It is appreciated that a combination of

purified and unpurified reagents - including enzyme(s) or substrates(s) or both – may be used.

It is appreciated that more than one source material may be used as a source of phytase activity. This is serviceable as one way to achieve a timed release of reagent(s) from source material(s), where release from different reagents from their source materials occur differentially, for example as injested source materials are digested *in vivo* or as source materials are processed in *in vitro* applications. The use of more than one source material of phytase activity is also serviceable to obtain phytase activities under a range of conditions and fluctuations thereof, that may be encountered - such as a range of pH values, temperatures, salinities, and time intervals – for example during different processing steps of an application. The use of different source materials is also serviceable in order to obtain different reagents, as exemplified by one or more forms or isomers of phytase and/or phytate and/or other materials.

It is appreciated that a single source material, such a trangenic plant species (or plant parts thereof), may be a source material of both phytase and phytate; and that enzymes and substrates may be differentially compartmentalized within said single source – *e.g.* secreted vs. non-secreted, differentially expressed and/or having differential abundances in different plant parts or organs or tissues or in subcellular compartments within the same plant part or organ or tissue. Purification of the phytase molecules contained therein may comprise isolating and/or further processing of one or more desirable plant parts or organs or tissues or subcellular compartments.

In a particular aspect, this invention provides a method of catalyzing *in vivo* and/or *in vitro* reactions using seeds containing enhanced amounts of enzymes. The method comprises adding transgenic, non-wild type seeds, preferably in a ground form, to a reaction mixture and allowing the enzymes in the seeds to increase the rate of reaction. By directly adding the seeds to the reaction mixture the method provides a solution to the more expensive and cumbersome process of extracting and purifying the enzyme. Methods of treatment are also provided whereby an organism lacking a sufficient supply of an enzyme is administered the enzyme in the form of seeds from one or more plant species, preferably transgenic plant species, containing enhanced

amounts of the enzyme. Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes U.S. Patent No. 5,543,576 (Van Ooijen *et al.*) and U.S. Patent No. 5,714,474 (Van Ooijen *et al.*), although these reference do not teach the inventive molecules of the instant application and instead teach the use of fungal phytases.

In a particular non-limiting aspect, the instant phytase molecules are serviceable for generating recombinant digestive system life forms (or microbes or flora) and for the administration of said recombinant digestive system life forms to animals. Administration may be optionally performed alone or in combination with other enzymes and/or with other life forms that can provide enzymatic activity in a digestive system, where said other enzymes and said life forms may be may recombinant or otherwise. For example, administration may be performed in combination with xylanolytic bacteria

In a non-limiting aspect, the present invention provides a method for steeping corn or sorghum kernels in warm water containing sulfur dioxide in the presence of an enzyme preparation comprising one or more phytin-degrading enzymes, preferably in such an amount that the phytin present in the corn or sorghum is substantially degraded. The enzyme preparation may comprise phytase and/or acid phosphatase and optionally other plant material degrading enzymes. The steeping time may be 12 to 18 hours. The steeping may be interrupted by an intermediate milling step, reducing the steeping time. In a preferred embodiment, corn or sorghum kernels are steeped in warm water containing sulfur dioxide in the presence of an enzyme preparation including one or more phytin-degrading enzymes, such as phytase and acid phosphatases, to eliminate or greatly reduce phytic acid and the salts of phytic acid. Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes U.S. Patent No. 4,914,029 (Caransa *et al.*) and EP 0321004 (Vaara *et al.*), although these reference do not teach the inventive molecules of the instant application.

In a non-limiting aspect, the present invention provides a method to obtain a bread dough having desirable physical properties such as non-tackiness and elasticity and a bread product of superior quality such as a specific volume comprising adding phytase molecules to the bread dough. In a preferred embodiment, phytase molecules of the instant invention are added to a working bread dough preparation that is subsequently formed and baked. Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes JP 03076529 (Hara *et al.*), although this reference does not teach the inventive phytase molecules of the instant application.

In a non-limiting aspect, the present invention provides a method to produce improved soybean foodstuffs. Soybeans are combined with phytase molecules of the instant invention to remove phytic acid from the soybeans, thus producing soybean foodstuffs that are improved in their supply of trace nutrients essential for consuming organisms and in its digestibility of proteins. In a preferred embodiment, in the production of soybean milk, phytase molecules of the instant invention are added to or brought into contact with soybeans in order to reduce the phytic acid content. In a non-limiting exemplification, the application process can be accelerated by agitating the soybean milk together with the enzyme under heating or by a conducting a mixing-type reaction in an agitation container using an immobilized enzyme. Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes JP 59166049 (Kamikubo *et al.*), although this reference does not teach the inventive molecules of the instant application.

In one aspect, the instant invention provides a method of producing an admixture product for drinking water or animal feed in fluid form, and which comprises using mineral mixtures and vitamin mixtures, and also novel phytase molecules of the instant invention. In a preferred embodiment, there is achieved a correctly dosed and composed mixture of necessary nutrients for the consuming organism without any risk of precipitation and destruction of important minerals/vitamins, while at the same time optimum utilization is made of the phytin-bound phosphate in the feed. Additional

details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes EP 0772978 (Bendixen *et al.*), although this reference does not teach the inventive molecules of the instant application.

It is appreciated that the phytase molecules of the instant invention may also be used to produce other alcoholic and non-alcoholic drinkable foodstuffs (or drinks) based on the use of molds and/or on grains and/or on other plants. These drinkable foodstuffs include liquors, wines, mixed alcoholic drinks (*e.g.* wine coolers, other alcoholic coffees such as Irish coffees, etc.), beers, near-beers, juices, extracts, homogenates, and purees. In a preferred exemplification, the instantly disclosed phytase molecules are used to generate transgenic versions of molds and/or grains and/or other plants serviceable for the production of such drinkable foodstuffs. In another preferred exemplification, the instantly disclosed phytase molecules are used as additional ingredients in the manufacturing process and/or in the final content of such drinkable foodstuffs. Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. However - due to the novelty of the instant invention - references in the publicly available literature do not teach the inventive molecules instantly disclosed.

In another non-limiting exemplification, the present invention provides a means to obtain refined sake having a reduced amount of phytin and an increased content of inositol. Such a sake may have – through direct and/or psychogenic effects - a preventive action on hepatic disease, arteriosclerosis, and other diseases. In a preferred embodiment, a sake is produced from rice Koji by multiplying a rice Koji mold having high phytase activity as a raw material. It is appreciated that the phytase molecules of the instant invention may be used to produce a serviceable mold with enhanced activity (preferably a transgenic mold) and/or added exogenously to augment the effects of a Koji mold. The strain is added to boiled rice and Koji is produced by a conventional procedure. In a preferred exemplification, the prepared Koji is used, the whole rice is prepared at two stages and Sake is produced at constant Sake temperature of 15°C to give the objective refined Sake having a reduced amount of phytin and an increased amount of inositol. Additional details regarding this approach are in the public

literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes JP 06153896 (Soga *et al.*) and JP 06070749 (Soga *et al.*), although these references do not teach the inventive molecules of the instant application

In a non-limiting aspect, the present invention provides a method to obtain an absorbefacient capable of promoting the absorption of minerals including ingested calcium without being digested by gastric juices or intestinal juices at a low cost. In a preferred embodiment, the mineral absorbefacient contains a partial hydrolysate of phytic acid as an active ingredient. Preferably, a partial hydrolyzate of the phytic acid is produced by hydrolyzing the phytic acid or its salts using novel phytase molecules of the instant invention. The treatment with the phytase molecules may occur either alone and/or in a combination treatment (to inhibit or to augment the final effect), and is followed by inhibiting the hydrolysis within a range so as not to liberate all the phosphate radicals. Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes JP 04270296 (Hoshino), although reference in the publicly available literature do not teach the inventive molecules of the instant application.

In a non-limiting aspect, the present invention provides a method (and products therefrom) to produce an enzyme composition having an additive or preferably a synergistic phytate hydrolyzing activity; said composition comprises novel phytase molecules of the instant invention and one or more additional reagents to achieve a composition that is serviceable for a combination treatment. In a preferred embodiment, the combination treatment of the present invention is achieved with the use of at least two phytases of different position specificity, *i.e.* any combinations of 1-, 2-, 3-, 4-, 5-, and 6-phytases. By combining phytases of different position specificity an additive or synergistic effect is obtained. Compositions such as food and feed or food and feed additives comprising such phytases in combination are also included in this invention as are processes for their preparation. Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes WO9

830681 (Ohmann *et al.*), although references in the publicly available literature do not teach the use of the inventive molecules of the instant application.

In another preferred embodiment, the combination treatment of the present invention is achieved with the use of an acid phosphatase having phytate hydrolyzing activity at a pH of 2.5, in a low ratio corresponding to a pH 2.5:5.0 activity profile of from about 0.1:1.0 to 10:1, preferably of from about 0.5:1.0 to 5:1, or from about 0.8:1.0 to 3:1, or from about 0.8:1.0 to 2:1. The enzyme composition preferably displays a higher synergetic phytate hydrolyzing efficiency through thermal treatment. The enzyme composition is serviceable in the treatment of foodstuffs (drinkable and solid food, feed and fodder products) to improve phytate hydrolysis. Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes U.S. Patent No. 5,554,399 (Vanderbeke *et al.*) and U.S. Patent No. 5,443,979 (Vanderbeke *et al.*) which rather teach the use of fungal (in particular *Aspegillus*) phytases.

In a non-limiting aspect, the present invention provides a method (and products therefrom) to produce composition comprised of the instant novel phytate-acting enzyme in combination with one or more additional enzymes that act on polysaccharides. Such polysaccharides can be selected from the group consisting of arabinans, fructans, fucans, galactans, galacturonans, glucans, mannans, xylans, levan, fucoidan, carrageenan, galactocarolose, pectin, pectic acid, amylose, pullulan, glycogen, amylopectin, cellulose, carboxylmethylcellulose, hydroxypropylmethylcellulose, dextran, pustulan, chitin, agarose, keratan, chondroitin, dermatan, hyaluronic acid, alginic acid, and polysaccharides containing at least one aldose, ketose, acid or amine selected from the group consisting of erythrose, threose, ribose, arabinose, xylose, lyxose, allose, altrose, glucose, mannose, gulose, idose, galactose, talose, erythrulose, ribulose, xylulose, psicose, fructose, sorbose, tagatose, glucuronic acid, gluconic acid, glucaric acid, galacturonic acid, mannuronic acid, glucosamine, galactosamine and neuraminic acid.

In a particular aspect, the present invention provides a method (and products therefrom) to produce composition having a synergistic phytate hydrolyzing activity comprising one or more novel phytase molecules of the instant invention, a cellulase (including preferably but not exclusively a xylanase), optionally a protease, and optionally one or more additonal reagents. In preferred embodiments, such combination treatments are serviceable in the treatment of foodstuffs, wood products, such as paper products, and as cleansing solutions and solids.

In one non-limiting exemplification, the instant phytase molecules are serviceable in combination with cellulosome components. It is known that cellulases of many cellulolytic bacteria are organized into discrete multienzyme complexes, called cellulosomes. The multiple subunits of cellulosomes are composed of numerous functional domains, which interact with each other and with the cellulosic substrate. One of these subunits comprises a distinctive new class of noncatalytic scaffolding polypeptide, which selectively integrates the various cellulase and xylanase subunits into the cohesive complex. Intelligent application of cellulosome hybrids and chimeric constructs of cellulosomal domains should enable better use of cellulosic biomass and may offer a wide range of novel applications in research, medicine and industry.

In another non-limiting exemplification, the instant phytase molecules are serviceable - either alone or in combination treatments - in areas of biopulping and biobleaching where a reduction in the use of environmentally harmful chemicals traditionally used in the pulp and paper industry is desired. Waste water treatment represents another vast application area where biological enzymes have been shown to be effective not only in colour removal but also in the bioconversion of potentially noxious substances into useful bioproducts.

In another non-limiting exemplification, the instant phytase molecules are serviceable for generating life forms that can provide at least one enzymatic activity - either alone or in combination treatments - in the treatment of digestive systems of organisms. Particularly relevant organisms to be treated include non-ruminant organisms. Specifically, it is appreciated that this approach may be performed alone or in combination with other biological molecules (for example, xylanases) to generate a

recombinant host that expresses a plurality of biological molecules. It is also appreciated that the administration of the instant phytase molecules and/or recombinant hosts expressing the instant phytase molecules may be performed either alone or in combination with other biological molecules, and/or life forms that can provide enzymatic activities in a digestive system - where said other enzymes and said life forms may be may recombinant or otherwise. For example, administration may be performed in combination with xylanolytic bacteria

For example, in addition to phytate, many organisms are also unable to adequately digest hemicelluloses. Hemicelluloses or xylans are major components (35%) of plant materials. For ruminant animals, about 50% of the dietary xylans are degraded, but only small amounts of xylans are degraded in the lower gut of nonruminant animals and humans. In the rumen, the major xylanolytic species are Butyrivibrio fibrisolvens and Bacteroides ruminicola. In the human colon, Bacteroides ovatus and Bacteroides fragilis subspecies "a" are major xylanolytic bacteria. Xylans are chemically complex, and their degradation requires multiple enzymes. Expression of these enzymes by gut bacteria varies greatly among species. Butyrivibrio fibrisolvens makes extracellular xylanases but Bacteroides species have cell-bound xylanase activity. Biochemical characterization of xylanolytic enzymes from gut bacteria has not been done completely. A xylosidase gene has been cloned from B. fibrosolvens 113. The data from DNA hybridizations using a xylanase gene cloned from B. fibrisolvens 49 indicate this gene may be present in other B. fibrisolvens strains. A cloned xylanase from Bact. ruminicola was transferred to and highly expressed in Bact. fragilis and Bact. uniformis. Arabinosidase and xylosidase genes from Bact. ovatus have been cloned and both activities appear to be catalyzed by a single, bifunctional, novel enzyme.

Accordingly, it is appreciated that the present phytase molecules are serviceable for 1) transferring into a suitable host (such as Bact. fragilis or Bact. uniformis); 2) achieving adequate expression in a resultant recombinant host; and 3) administering said recombinant host to organisms to improve the ability of the treated organisms to degrade phytate. Continued research in genetic and biochemical areas will provide

knowledge and insights for manipulation of digestion at the gut level and improved understanding of colonic fiber digestion.

Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes U.S. Patent No. 5,624,678 (Bedford *et al.*), U.S. Patent No. 5,683,911 (Bodie *et al.*), U.S. Patent No. 5,720,971 (Beauchemin *et al.*), U.S. Patent No. 5,759,840 (Sung *et al.*), U.S. Patent No. 5,770,012 (Cooper), U.S. Patent No. 5,786,316 (Baeck *et al.*), U.S. Patent No. 5,817,500 (Hansen *et al.*), and journal articles (Jeffries, 1996; Prade, 1996; Bayer *et al.*, 1994; Duarte *et al.*, 1994; Hespell and Whitehead, 1990; Wong *et al.*, 1988), although these reference do not teach the inventive phytase molecules of the instant application, nor do they all teach the addition of phytase molecules in the production of foodstuffs, wood products, such as paper products, and as cleansing solutions and solids. In contrast, the instant invention teaches that phytase molecules – preferably the inventive phytase molecules of the instant application - may be added to the reagent(s) disclosed in order to obtain preparations having an additional phytase activity. Preferably, said reagent(s) the additional phytase molecules and will not inhibit each other, more preferably said reagent(s) the additional phytase molecules will have an overall additive effect, and more preferably still said reagent(s) the additional phytase molecules will have an overall synergistic effect.

In a non-limiting aspect, the present invention provides a method (and products therefrom) for enhancement of phytate phosphorus utilization and treatment and prevention of tibial dyschondroplasia in animals, particularly poultry, by administering to animals a feed composition containing a hydroxylated vitamin D_3 derivative. The vitamin D_3 derivative is preferably administered to animals in feed containing reduced levels of calcium and phosphorus for enhancement of phytate phosphorus utilization. Accordingly, the vitamin D_3 derivative is preferably administered in combination with novel phytase molecules of the instant invention for further enhancement of phytate phosphorus utilization. Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes U.S. Patent No. 5,516,525

(Edwards *et al.*) and U.S. Patent No. 5,366,736 (Edwards *et al.*), U.S. Patent No. 5,316,770 (Edwards *et al.*) although these reference do not teach the inventive molecules of the instant application.

In a non-limiting aspect, the present invention provides a method (and products therefrom) to obtain foodstuff that 1) comprises phytin that is easily absorbed and utilized in a form of inositol in a body of an organism; 2) that is capable of reducing phosphorus in excrementary matter; and 3) that is accordingly useful for improving environmental pollution. Said foodstuff is comprised of an admixture of a phytin-containing grain, a lactic acid-producing microorganism, and a novel phytase molecule of the instant invention. In a preferred embodiment, said foodstuff is produced by compounding a phytin- containing grain (preferably, *e.g.* rice bran) with an effective microbial group having an acidophilic property, producing lactic acid, without producing butyric acid, free from pathogenicity, and a phytase. Examples of an effective microbial group include *e.g. Streptomyces sp.* (American Type Culture Collection No. ATCC 3004) belonging to the group of actinomyces and *Lactobacillus sp.* (IFO 3070) belonging to the group of lactobacilli. Further, a preferable amount of addition of an effective microbial group is 0.2 wt. % in terms of bacterial body weight based on a grain material. Furthermore, the amount of the addition of the phytase is preferably 1-2 wt. % based on the phytin in the grain material. Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes JP 08205785 (Akahori *et al.*), although references in the publicly available literature do not teach the inventive molecules of the instant application.

In a non-limiting aspect, the present invention provides a method for improving the solubility of vegetable proteins. More specifically, the invention relates to methods for the solubilization of proteins in vegetable protein sources, which methods comprise treating the vegetable protein source with an efficient amount of one or more phytase enzymes – including phytase molecules of the instant invention - and treating the vegetable protein source with an efficient amount of one or more proteolytic enzymes. In another aspect, the invention provides animal feed additives comprising a phytase and one or more proteolytic enzymes. Additional details regarding this approach are in

the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes EP 0756457 (WO 9528850 A1) (Nielsen and Knap), although references in the publicly available literature do not teach the inventive molecules of the instant application.

In a non-limiting aspect, the present invention provides a method of producing a plant protein preparation comprising dispersing vegetable protein source materials in water at a pH in the range of 2 to 6 and admixing phytase molecules of the instant invention therein. The acidic extract containing soluble protein is separated and dried to yield a solid protein of desirable character. One or more proteases can also be used to improve the characteristics of the protein. Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes U.S. Patent No. 3,966,971 (Morehouse *et al.*), although references in the publicly available literature do not teach the inventive molecules of the instant application.

In a non-limiting aspect, the present invention provides a method (and products thereof) to activate inert phosphorus in soil and/or compost, to improve the utilization rate of a nitrogen compound, and to suppress propagation of pathogenic molds by adding three reagents, phytase, saponin and chitosan, to the compost. In a non-limiting embodiment the method can comprise treating the compost by 1) adding phytase-containing microorganisms in media – preferably recombinant hosts that overexpress the novel phytase molecules of the instant invention - *e.g.* at 100 ml media/100 kg wet compost; 2) alternatively also adding a phytase-containing plant source - such as wheat bran - *e.g.* at 0.2 to 1 kg/100 kg wet compost; 3) adding a saponin-containing source - such as peat, mugworts and yucca plants – *e.g.* at 0.5 to 3.0g/kg ; 4) adding chitosan-containing materials – such as pulverized shells of shrimps, crabs, etc. – *e.g.* at 100 to 300g/kg wet compost. In another non-limiting embodiment, recombinant sources the three reagents, phytase, saponin, and chitosan, are used. Additional details regarding this approach are in the public literature and/or are known to the skilled artisan. In a particular non-limiting exemplification, such publicly available literature includes JP 07277865 (Toya Taisuke), although references in the publicly available literature do not teach the inventive molecules of the instant application.

Fragments of the full length gene of the present invention may be used as a hybridization probe for a cDNA or a genomic library to isolate the full length DNA and to isolate other DNAs which have a high sequence similarity to the gene or similar biological activity. Probes of this type have at least 10, preferably at least 15, and even more preferably at least 30 bases and may contain, for example, at least 50 or more bases. The probe may also be used to identify a DNA clone corresponding to a full length transcript and a genomic clone or clones that contain the complete gene including regulatory and promotor regions, exons, and introns.

In another embodiment, transgenic non-human organisms are provided which contain a heterolgous sequence encoding a phytase of the invention (e.g., SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14). Various methods to make the transgenic animals of the subject invention can be employed. Generally speaking, three such methods may be employed. In one such method, an embryo at the pronuclear stage (a "one cell embryo") is harvested from a female and the transgene is microinjected into the embryo, in which case the transgene will be chromosomally integrated into both the germ cells and somatic cells of the resulting mature animal. In another such method, embryonic stem cells are isolated and the transgene incorporated therein by electroporation, plasmid transfection or microinjection, followed by reintroduction of the stem cells into the embryo where they colonize and contribute to the germ line. Methods for microinjection of mammalian species is described in U.S. Pat. No. 4,873,191. In yet another such method, embryonic cells are infected with a retrovirus containing the transgene whereby the germ cells of the embryo have the transgene chromosomally integrated therein. When the animals to be made transgenic are avian, because avian fertilized ova generally go through cell division for the first twenty hours in the oviduct, microinjection into the pronucleus of the fertilized egg is problematic due to the inaccessibility of the pronucleus. Therefore, of the methods to make transgenic animals described generally above, retrovirus infection is preferred for avian species, for example as described in U.S. Pat No. 5,162,215. If micro-injection is to be used with avian species, however, a published procedure by Love *et al.*, (*Biotechnol.*, 12, Jan 1994) can be utilized whereby the embryo is obtained from a sacrificed hen

approximately two and one-half hours after the laying of the previous laid egg, the transgene is microinjected into the cytoplasm of the germinal disc and the embryo is cultured in a host shell until maturity. When the animals to be made transgenic are bovine or porcine, microinjection can be hampered by the opacity of the ova thereby making the nuclei difficult to identify by traditional differential interference-contrast microscopy. To overcome this problem, the ova can first be centrifuged to segregate the pronuclei for better visualization.

The "non-human animals" of the invention bovine, porcine, ovine and avian animals (*e.g.*, cow, pig, sheep, chicken). The "transgenic non-human animals" of the invention are produced by introducing "transgenes" into the germline of the non-human animal. Embryonal target cells at various developmental stages can be used to introduce transgenes. Different methods are used depending on the stage of development of the embryonal target cell. The zygote is the best target for micro-injection. The use of zygotes as is target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (Brinster *et al.*, *Proc. Natl. Acad. Sci.* USA 82:4438-4442, 1985). As a consequence, all cells of the transgenic non-human animal will carry the incorporated transgene. This will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene.

The term "transgenic" is used to describe an animal which includes exogenous genetic material within all of its cells. A "transgenic" animal can be produced by cross-breeding two chimeric animals which include exogenous genetic material within cells used in reproduction. Twenty-five percent of the resulting offspring will be transgenic *i.e.*, animals which include the exogenous genetic material within all of their cells in both alleles, 50% of the resulting animals will include the exogenous genetic material within one allele and 25% will include no exogenous genetic material.

In the microinjection method useful in the practice of the subject invention, the transgene is digested and purified free from any vector DNA, *e.g.*, by gel electrophoresis. It is preferred that the transgene include an operatively associated promoter which interacts with cellular proteins involved in transcription, ultimately

resulting in constitutive expression. Promoters useful in this regard include those from cytomegalovirus (CMV) , Moloney leukemia virus (MLV), and herpes virus, as well as those from the genes encoding metallothionin, skeletal actin, P-enolpyruvate carboxylase (PEPCK), phosphoglycerate (PGK), DHFR, and thymidine kinase. Promoters for viral long terminal repeats (LTRs) such as Rous Sarcoma Virus can also be employed. When the animals to be made transgenic are avian, preferred promoters include those for the chicken ϑ-globin gene, chicken lysozyme gene, and avian leukosis virus. Constructs useful in plasmid transfection of embryonic stem cells will employ additional regulatory elements well known in the art such as enhancer elements to stimulate transcription, splice acceptors, termination and polyadenylation signals, and ribosome binding sites to permit translation.

Retroviral infection can also be used to introduce transgene into a non-human animal, as described above. The developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Jaenich, R., *Proc. Natl. Acad. Sci.* USA 73:1260-1264, 1976). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (Hogan, *et al.* (1986) in Manipulating the Mouse Embryo, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.). The viral vector system used to introduce the transgene is typically a replication-defective retro virus carrying the transgene (Jahner, *et al.*, *Proc. Natl. Acad. Sci.* USA 82: 6927-6931, 1985; Van der Putten, *et al.*, *Proc. Natl. Acad. Sci.* USA 82: 6148-6152, 1985). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Van der Putten, *supra*; Stewart, *et al.*, *EMBO J.* 6: 383-388, 1987). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (D. Jahner *et al.*, *Nature* 298: 623-628, 1982). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells which formed the transgenic nonhuman animal. Further, the founder may contain various retro viral insertions of the transgene at different positions in the genome which generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line, albeit with low efficiency, by intrauterine retroviral infection of the midgestation embryo (D. Jahner *et al.*, *supra*).

A third type of target cell for transgene introduction is the embryonal stem cell (ES). ES cells are obtained from pre-implantation embryos cultured *in vitro* and fused with embryos (M. J. Evans *et al., Nature* 292:154-156, 1981; M. O. Bradley *et al., Nature* 309:255-258, 1984; Gossler, *et al., Proc. Natl. Acad. Sci.* USA 83:9065-9069, 1986; and Robertson *et al., Nature* 322:445-448, 1986). Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retro virus-mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a nonhuman animal. The ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal. (For review see Jaenisch, R., *Science* 240:1468-1474, 1988).

"Transformed" means a cell into which (or into an ancestor of which) has been introduced, by means of recombinant nucleic acid techniques, a heterologous nucleic acid molecule. "Heterologous" refers to a nucleic acid sequence that either originates from another species or is modified from either its original form or the form primarily expressed in the cell.

"Transgene" means any piece of DNA which is inserted by artifice into a cell, and becomes part of the genome of the organism (*i.e.*, either stably integrated or as a stable extrachromosomal element) which develops from that cell. Such a transgene may include a gene which is partly or entirely heterologous (*i.e.*, foreign) to the transgenic organism, or may represent a gene homologous to an endogenous gene of the organism. Included within this definition is a transgene created by the providing of an RNA sequence which is transcribed into DNA and then incorporated into the genome. The transgenes of the invention include DNA sequences which encode phytases or polypeptides having phytase activity, and include polynucleotides, which may be expressed in a transgenic non-human animal. The term "transgenic" as used herein additionally includes any organism whose genome has been altered by *in vitro* manipulation of the early embryo or fertilized egg or by any transgenic technology to induce a specific gene knockout. The term "gene knockout" as used herein, refers to the targeted disruption of a gene *in vivo* with complete loss of function that has been achieved by any transgenic technology familiar to those in the art. In one embodiment, transgenic animals having gene knockouts are those in which the target gene has been

rendered nonfunctional by an insertion targeted to the gene to be rendered non-functional by homologous recombination. As used herein, the term "transgenic" includes any transgenic technology familiar to those in the art which can produce an organism carrying an introduced transgene or one in which an endogenous gene has been rendered non-functional or "knocked out."

The transgene to be used in the practice of the subject invention is a DNA sequence comprising a sequence coding for a phytase or a polypeptide having phytase activity. In a one embodiment, a polynucleotide having a sequence as set forth in SEQ ID NO:1 or 3 or a sequence encoding a polypeptide having a sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14 is the transgene as the term is defined herein. Where appropriate, DNA sequences that encode proteins having phytase activity but differ in nucleic acid sequence due to the degeneracy of the genetic code may also be used herein, as may truncated forms, allelic variants and interspecies homologues.

After an embryo has been microinjected, colonized with transfected embryonic stem cells or infected with a retrovirus containing the transgene (except for practice of the subject invention in avian species which is addressed elsewhere herein) the embryo is implanted into the oviduct of a pseudopregnant female. The consequent progeny are tested for incorporation of the transgene by Southern blot analysis of blood or tissue samples using transgene specific probes. PCR is particularly useful in this regard. Positive progeny (G0) are crossbred to produce offspring (G1) which are analyzed for transgene expression by Northern blot analysis of tissue samples.

Thus, the present invention includes methods for increasing the phosphorous uptake in the transgenic animal and/or decreasing the amount of polltant in the manure of the transgenic organism by about 15%, about 20%, or about 20%, to about 50%.

The animals contemplated for use in the practice of the subject invention are those animals generally regarded as domesticated animals including pets (*e.g.*, canines, felines, avian species *etc.*) and those useful for the processing of food stuffs, *i.e.*, avian such as meat bred and egg laying chicken and turkey, ovine such as lamb, bovine such

as beef cattle and milk cows, piscine and porcine. For purposes of the subject invention, these animals are referred to as "transgenic" when such animal has had a heterologous DNA sequence, or one or more additional DNA sequences normally endogenous to the animal (collectively referred to herein as "transgenes") chromosomally integrated into the germ cells of the animal. The transgenic animal (including its progeny) will also have the transgene fortuitously integrated into the chromosomes of somatic cells.

In some instances it may be advantageous to deliver and express a phytase sequence of the invention locally (*e.g.*, within a particular tissue or cell type). For example, local expression of a phytase or digestive enzyme in the gut of an animal will assist in the digestion and uptake of, for example, phytate and phosporous, respectively. The nucleic sequence may be directly delivered to the salivary glands, tissue and cells and/or to the epithelial cells lining the gut, for example. Such delivery methods are known in the art and include electroporation, viral vectors and direct DNA uptake. Any polypeptide having phytase activity can be utilized in the methods of the invention (*e.g.*, those specficially described under this subsection 6.3.18, as well as those described in other sections of the invention).

For example, a nucleic acid constructs of the present invention will comprise nucleic acid molecules in a form suitable for uptake into target cells within a host tissue. The nucleic acids may be in the form of bare DNA or RNA molecules, where the molecules may comprise one or more structural genes, one or more regulatory genes, antisense strands, strands capable of triplex formation, or the like. Commonly, the nucleic acid construct will include at least one structural gene under the transcriptional and translational control of a suitable regulatory region. More usually, nucleic acid constructs of the present invention will comprise nucleic acids incorporated in a delivery vehicle to improve transfection efficiency, wherein the delivery vehicle will be dispersed within larger particles comprising a dried hydrophilic excipient material.

One such delivery vehicles comprises viral vectors, such as retroviruses, adenoviruses, and adeno-associated viruses, which have been inactivated to prevent self-replication but which maintain the native viral ability to bind a target host cell, deliver genetic material into the cytoplasm of the target host cell, and promote

expression of structural or other genes which have been incorporated in the particle. Suitable retrovirus vectors for mediated gene transfer are described in Kahn *et al.* (1992) *Circ. Res.* 71:1508-1517, the disclosure of which is incorporated herein by reference. A suitable adenovirus gene delivery is described in Rosenfeld *et al.* (1991) *Science* 252:431-434, the disclosure of which is incorporated herein by reference. Both retroviral and adenovirus delivery systems are described in Friedman (1989) *Science* 244:1275-1281, the disclosure of which is also incorporated herein by reference.

A second type of nucleic acid delivery vehicle comprises liposomal transfection vesicles, including both anionic and cationic liposomal constructs. The use of anionic liposomes requires that the nucleic acids be entrapped within the liposome. Cationic liposomes do not require nucleic acid entrapment and instead may be formed by simple mixing of the nucleic acids and liposomes. The cationic liposomes avidly bind to the negatively charged nucleic acid molecules, including both DNA and RNA, to yield complexes which give reasonable transfection efficiency in many cell types. See, Farhood *et al.* (1992) *Biochem. Biophys. Acta.* 1111:239-246, the disclosure of which is incorporated herein by reference. A particularly preferred material for forming liposomal vesicles is lipofectin which is composed of an equimolar mixture of dioleylphosphatidyl ethanolamine (DOPE) and dioleyloxypropyl-triethylammonium (DOTMA), as described in Felgner and Ringold (1989) *Nature* 337:387-388, the disclosure of which is incorporated herein by reference.

It is also possible to combine these two types of delivery systems. For example, Kahn *et al.* (1992), *supra.*, teaches that a retrovirus vector may be combined in a cationic DEAE-dextran vesicle to further enhance transformation efficiency. It is also possible to incorporate nuclear proteins into viral and/or liposomal delivery vesicles to even further improve transfection efficiencies. See, Kaneda *et al.* (1989) *Science* 243:375-378, the disclosure of which is incorporated herein by reference.

In another embodiment, a digestive aid containing an enzyme either as the sole active ingredient or in combination with one or more other agents and/or enzymes is provided (as described in co-pending application U.S. Serial No. _____, entitled "Dietary Aids and Methods of Use Thereof," filed May 25, 2000, the disclosure of

which is incorporated herein by reference in its entirety). The use of enzymes and other agents in digestive aids of livestock or domesticated animals not only improves the animal's health and life expectancy but also assists in increasing the health of livestock and in the production of foodstuffs from livestock.

Currently, some types of feed for livestock (*e.g.*, certain poultry feed) are highly supplemented with numerous minerals (*e.g.*, inorganic phosphorous), enzymes, growth factors, drugs, and other agents for delivery to the livestock. These supplements replace many of the calories and natural nutrients present in grain, for example.

By reducing or eliminating the inorganic phosphorous supplement and other supplements (*e.g.*, trace mineral salts, growth factors, enzymes, antibiotics) from the feed itself, the feed is able to carry more nutrient and energy. Accordingly, the remaining diet would contain more usable energy. For example, grain-oilseed meal diets generally contain about 3,200 kcal metabolizable energy per kilogram of diet, and mineral salts supply no metabolizable energy. Removal of the unneeded minerals and substitution with grain therefore increase the usable energy in the diet. Thus, the invention is differentiated over commonly used phytase containing feed. For example, in one embodiment, a biocompatible material is used that is resistant to digestion by the gastrointestinal tract of an organism.

In many organisms, including, for example, poultry or birds such as, for example, chickens, turkeys, geese, ducks, parrots, peacocks, ostriches, pheasants, quail, pigeons, emu, kiwi, loons, cockatiel, cockatoo, canaries, penguins, flamingoes, and dove, the digestive tract includes a gizzard which stores and uses hard biocompatible objects (*e.g.*, rocks and shells from shell fish) to help in the digestion of seeds or other feed consumed by a bird. A typical digestive tract of this general family of organisms, includes the esophagus which contains a pouch, called a crop, where food is stored for a brief period of time. From the crop, food moves down into the true stomach, or *proventriculus*, where hydrochloric acid and pepsin starts the process of digestion. Next, food moves into the gizzard, which is oval shaped and thick walled with powerful muscles. The chief function of the gizzard is to grind or crush food particles - a process which is aided by the bird swallowing small amounts of fine gravel or grit. From the

gizzard, food moves into the duodenum. The small intestine of birds is similar to mammals. There are two blind pouches or *ceca*, about 4-6 inches in length at the junction of the small and large intestine. The large intestine is short, consisting mostly of the rectum about 3-4 inches in length. The rectum empties into the *cloaca* and feces are excreted through the vent.

Hard, biocompatible objects consumed (or otherwise introduced) and presented in the gizzard provide a useful vector for delivery of various enzymatic, chemical, therapeutic and antibiotic agents. These hard substances have a life span of a few hours to a few days and are passed after a period of time. Accordingly, the invention provides coated, impregnated (*e.g.*, impregnated matrix and membranes) modified dietary aids for delivery of useful digestive or therapeutic agents to an organism. Such dietary aids include objects which are typically ingested by an organism to assist in digestion within the gizzard (*e.g.*, rocks or grit). The invention provides biocompatible objects that have coated thereon or impregnated therein agents useful as a digestive aid for an organism or for the delivery of a therapeutic or medicinal agent or chemical.

In a one embodiment, the invention provides a dietary aid, having a biocompatible composition designed for release of an agent that assists in digestion, wherein the biocompatible composition is designed for oral consumption and release in the digestive tract (*e.g.*, the gizzard) of an organism. "Biocompatible" means that the substance, upon contact with a host organism (*e.g.*, a bird), does not elicit a detrimental response sufficient to result in the rejection of the substance or to render the substance inoperable. Such inoperability may occur, for example, by formation of a fibrotic structure around the substance limiting diffusion of impregnated agents to the host organism therein or a substance which results in an increase in mortality or morbidity in the organism due to toxicity or infection. A biocompatible substance may be non-biodegradable or biodegradable. In one embodiment, the biocompatible composition is resistant to degradation or digestion by the gastrointestinal tract. In another embodiment, the biocompatible composition has the consistency of a rock or stone.

A non-biodegradable material useful in the invention is one that allows attachment or impregnation of a dietary agent. Such non-limiting non-biodegradable

materials include, for example, thermoplastics, such as acrylic, modacrylic, polyamide, polycarbonate, polyester, polyethylene, polypropylene, polystyrene, polysulfone, polyethersulfone, and polyvinylidene fluoride. Elastomers are also useful materials and include, for example, polyamide, polyester, polyethylene, polypropylene, polystyrene, polyurethane, polyvinyl alcohol and silicone (*e.g.*, silicone based or containing silica). The invention provides that the biocompatible composition can contain a plurality of such materials, which can be, *e.g.*, admixed or layered to form blends, copolymers or combinations thereof.

As used herein, a "biodegradable" material means that the composition will erode or degrade *in vivo* to form smaller chemical species. Degradation may occur, for example, by enzymatic, chemical or physical processes. Suitable biodegradable materials contemplated for use in the invention include, but are not limited to, poly(lactide)s, poly(glycolide)s, poly(lactic acid)s, poly(glycolic acid)s, polyanhydrides, polyorthoesters, polyetheresters, polycaprolactone, polyesteramides, polycarbonate, polycyanoacrylate, polyurethanes, polyacrylate, and the like. Such materials can be admixed or layered to form blends, copolymers or combinations thereof.

It is contemplated that a number different biocompatible substances may be ingested or otherwise provided to the same organism simultaneously, or in various combinations (*e.g.*, one material before the other). In addition, the biocompatible substance may be designed for slow passage through the digestive tract. For example, large or fatty substances tend to move more slowly through the digestive tract, accordingly, a biocompatible material having a large size to prevent rapid passing in the digestive tract can be used. Such large substances can be a combination of non-biodegradable and biodegradable substances. For example, a small non-biodegradable substance can be encompassed by a biodegradable substance such that over a period of time the biodegradable portion will be degraded allowing the non-biodegradable portion to pass through the digestive trace. In addition, it is recognized that any number of flavorings can be provided to the biocompatible substance to assist in consumption.

Any number of agents alone or in combination with other agents can be coated on the biocompatible substance including polypeptides (*e.g.*, enzymes, antibodies,

cytokines or therapeutic small molecules), and antibiotics, for example. Examples of particular useful agents are listed in Table 1 and 2, below. It is also contemplated that cells can be encapsulated into the biocompatible material of the invention and used to deliver the enzymes or therapeutics. For example, porous substances can be designed that have pores large enough for cells to grow in and through and that these porous materials can then be taken into the digestive tract. For example, the biocompatible substance can be comprised of a plurality of microfloral environments (*e.g.*, different porosity, pH etc.) that provide support for a plurality of cell types. The cells can be genetically engineered to deliver a particular drug, enzyme or chemical to the organism. The cells can be eukaryotic or prokaryotic.

Table 1

Treatment Class	Chemical	Description
Antibiotics	Amoxycillin and Its Combination Mastox Injection (Amoxycillin and Cloxacillin)	Treatment Against Bacterial Diseases Caused By Gram + and Gram – Bacteria
	Ampicillin and Its Combination Biolox Injection (Ampicillin and Cloxacillin)	Treatment Against Bacterial Diseases Caused By Gram + And Gram – Bacteria.
	Nitrofurazone + Urea Nefrea Bolus	Treatment Of Genital Infections
	Trimethoprim + Sulphamethoxazole Trizol Bolus	Treatment Of Respiratory Tract Infections, Gastro Intestinal Tract Infections, Urino- Genital Infections.
	Metronidazole and Furazolidone Metofur Bolus	Treatment Of Bacterial And Protozoal Diseases.
	Phthalylsulphathiazole, Pectin and Kaolin Pectolin Bolus Suspension	Treatment Of Bacterial And Non-Specific Diarrhoea, Bacillary Dysentry And Calf Scours.
Antihelmintics	Ectoparasiticide Germex Ointment (Gamma Benzene Hexachloride, Proflavin Hemisulphate and Cetrimide)	Ectoparasiticide and Antiseptic
	Endoparasiticides > Albendazole and Its Combination Alben (Albendazole) Suspension (Albendazole 2.5%) Plus Suspension (Albendazole 5%) Forte Bolus (Albendazole 1.5 Gm.) Tablet (Albendazole 600 Mg.) Powder(Albendazole 5%, 15%)	Prevention And Treatment Of Roundworm, Tapeworm and Fluke Infestations
	Alpraz (Albendazole and Praziquantel)Tablet	Prevention And Treatment Of Roundworm and Tapeworm Infestation In Canines and Felines.
	Oxyclozanide and Its Combination Clozan (Oxyclozanide) Bolus, Suspension	Prevention and Treatment Of Fluke Infestations
	Tetzan (Oxyclozanide and Tetramisole Hcl) Bolus, Suspension	Prevention and Treatment Of Roundworm and Fluke Infestations
	Fluzan (Oxyclozanide and Levamisole Hcl) Bolus, Suspension	Prevention and Treatment Of Roundworm Infestations and Increasing Immunity
	Levamisole Nemasol Injection Wormnil Powder	Prevention and Treatment Of Roundworm Infestations and Increasing Immunity.

	Fenbendazole Fenzole Tablet (Fenbendazole150 Mg.) Bolus (Fenbendazole 1.5 Gm.) Powder (Fenbendazole 2.5% W/W)	Prevention And Treatment of Roundworm and Tapeworm Infestations
Tonics	Vitamin B Complex, Amino Acids and Liver Extract Heptogen Injection	Treatment Of Anorexia, Hepatitis, Debility, Neuralgic Convulsions Emaciation and Stunted Growth.
	Calcium Levulinate With Vit.B_{12} and Vit D_3 Hylactin Injection	Prevention and treatment of hypocalcaemia, supportive therapy in sick conditions (especially hypothermia) and treatment of early stages of rickets.
Animal Feed Supplements	Essential Minerals, Selenium and Vitamin E Gynolactin Bolus	Treatment Of Anoestrus Causing Infertility and Repeat Breeding In Dairy Animals and Horses.
	Essential Minerals, Vitamin E, and Iodine Hylactin Powder	Infertility, Improper Lactation, Decreased Immunity, Stunted Growth and Debility.
	Essential Electrolytes With Vitamin C Electra – C Powder	Diarrhoea, Dehydration, Prior to and after Transportation, In Extreme temperatures (High Or Low) and other Conditions of stress.
	Pyrenox Plus (Diclofenac Sodium + Paracetamol) Bolus, Injection.	Treatment Of Mastitis, Pyrexia Post Surgical Pain and Inflammation, Prolapse Of Uterus, Lameness and Arthritis.

Table 2. Therapeutic Formulations

Product	Description
Acutrim® (phenylpropanolamine)	Once-daily appetite suppressant tablets.
The Baxter® Infusor	For controlled intravenous delivery of anticoagulants, antibiotics, chemotherapeutic agents, and other widely used drugs.
Catapres-TTS® (clonidine transdermal therapeutic system)	Once-weekly transdermal system for the treatment of hypertension.
Covera HS3 (verapamil hydrochloride)	Once-daily Controlled-Onset Extended-Release (COER-24) tablets for the treatment of hypertension and angina pectoris.
DynaCirc CR® (isradipine)	Once-daily extended release tablets for the treatment of hypertension.
Efidac 24® (chlorpheniramine maleate)	Once-daily extended release tablets for the relief of allergy symptoms.
Estraderm® (estradiol transdermal system)	Twice-weekly transdermal system for treating certain postmenopausal symptoms and preventing osteoporosis
Glucotrol XL® (glipizide)	Once-daily extended release tablets used as an adjunct to diet for the control of hyperglycemia in patients with non-insulin-dependent diabetes mellitus.
IVOMEC SR® Bolus (ivermectin)	Ruminal delivery system for season-long control of major internal and external parasites in cattle.
Minipress XL® (prazosin)	Once-daily extended release tablets for the treatment of hypertension.
NicoDerm® CQ™ (nicotine transdermal system)	Transdermal system used as a once-daily aid to smoking cessation for relief of nicotine withdrawal symptoms.
Procardia XL® (nifedipine)	Once-daily extended release tablets for the treatment of angina and hypertension.
Sudafed® 24 Hour (pseudoephedrine)	Once-daily nasal decongestant for relief of colds, sinusitis, hay fever and other respiratory allergies.
Transderm-Nitro® (nitroglycerin transdermal system)	Once-daily transdermal system for the prevention of angina pectoris due to coronary artery disease.
Transderm Scop® (scopolamin transdermal system)	Transdermal system for the prevention of nausea and vomiting associated with motion sickness.
Volmax (albuterol)	Extended release tablets for relief of bronchospasm in patients with reversible obstructive airway disease.
Actisite®	(tetracycline hydrochloride) Periodontal fiber used as an adjunct to scaling and root planing for reduction of pocket depth and bleeding on probing in patients with adult periodontitis.
ALZET®	Osmotic pumps for laboratory research.
Amphotec® (amphotericin B cholesteryl sulfate complex for injection)	AMPHOTEC® is a fungicidal treatment for invasive aspergillosis in patients where renal impairment or unacceptable toxicity precludes use of amphotericin B in effective doses and in patients with invasive aspergillosis where prior amphotericin B therapy has failed.
BiCitra® (sodium citrate and citric acid)	Alkalinizing agent used in those conditions where long-term maintenance of alkaline urine is desirable.
Ditropan® (oxybutynin chloride)	For the relief of symptoms of bladder instability associated with uninhibited neurogenic or reflex neurogenic bladder (*i.e.*, urgency, frequency, urinary leakage, urge incontinence, dysuria).
Ditropan® XL (oxybutynin chloride)	is a once-daily controlled-release tablet indicated for the treatment of overactive bladder with symptoms of urge urinary incontinence, urgency and frequency.
DOXIL® (doxorubicin HCl liposome injection)	
Duragesic® (fentanyl transdermal system) CII	72-hour transdermal system for management of chronic pain in patients who require continuous opioid analgesia for pain that cannot be managed by lesser means such as acetaminophen-opioid combinations, non-steroidal analgesics, or PRN dosing with short-acting opioids.

Elmiron® (pentosan polysulfate sodium)	Indicated for the relief of bladder pain or discomfort associated with interstitial cystitis.
ENACT AirWatch™	An asthma monitoring and management system.
Ethyol® (amifostine)	Indicated to reduce the cumulative renal toxicity associated with repeated administration of cisplatin in patients with advanced ovarian cancer or non-small cell lung cancer. Indicated to reduce the incidence of moderate to severe xerostomia in patients undergoing post-operative radiation treatment for head and neck cancer, where the radiation port includes a substantial portion of the parotid glands.
Mycelex® Troche (clotrimazole)	For the local treatment of oropharyngeal candidiasis. Also indicated prophylactically to reduce the incidence of oropharyngeal candidiasis in patients immunocompromised by conditions that include chemotherapy, radiotherapy, or steroid therapy utilized in the treatment of leukemia, solid tumors, or renal transplantation.
Neutra-Phos® (potassium and sodium phosphate)	a dietary/nutritional supplement
PolyCitra® -K Oral Solution and PolyCitra® -K Crystals (potassium citrate and citric acid)	Alkalinizing agent useful in those conditions where long-term maintenance of an alkaline urine is desirable, such as in patents with uric acid and cystine calculi of the urinary tract, especially when the administration of sodium salts is undesirable or contraindicated
PolyCitra® -K Syrup and LC (tricitrates)	Alkalinizing agent useful in those conditions where long-term maintenance of an alkaline urine is desirable, such as in patients with uric acid and cystine calculi of the urinary tract.
Progestasert® (progesterone)	Intrauterine Progesterone Contraceptive System
Testoderm® Testoderm® with Adhesive and Testoderm® TTS CIII	Testosterone Transdermal System The Testoderm® products are indicated for replacement therapy in males for conditions associated with a deficiency or absence of endogenous testosterone: (1) Primary hypogonadism (congenital or acquired) or (2) Hypogonadotropic hypogonadism (congenital or acquired).
Viadur™ (leuprolide acetate implant)	Once-yearly implant for the palliative treatment of prostate cancer

Certain agents can be designed to become active or in activated under certain conditions (*e.g.*, at certain pH's, in the presence of an activating agent etc.). In addition, it may be advantageous to use pro-enzymes in the compositions of the invention. For example, a pro-enzymes can be activated by a protease (*e.g.*, a salivary protease that is present in the digestive tract or is artificially introduced into the digestive tract of an organism). It is contemplated that the agents delivered by the biocompatible compositions of the invention are activated or inactivated by the addition of an activating agent which may be ingested by, or otherwise delivered to, the organism. Another mechanism for control of the agent in the digestive tract is an environment sensitive agent that is activated in the proper digestive compartment. For example, an agent may be inactive at low pH but active at neutral pH. Accordingly, the agent would be inactive in the gut but active in the intestinal tract. Alternatively, the agent can become active in response to the presence of a microorganism specific factor (*e.g.*, microorganisms present in the intestine).

Accordingly, the potential benefits of the present invention include, for example, (1) reduction in or possible elimination of the need for mineral supplements (*e.g.*, inorganic phosphorous supplements), enzymes, or therapeutic drugs for animal (including fish) from the daily feed or grain thereby increasing the amount of calories and nutrients present in the feed, and (2) increased health and growth of domestic and non-domestic animals including, for example, poultry, porcine, bovine, equine, canine, and feline animals.

A large number of enzymes can be used in the methods and compositions of the present invention in addition to the phytases of the invention. These enzymes include enzymes necessary for proper digestion of consumed foods, or for proper metabolism, activation or derivation of chemicals, prodrugs or other agents or compounds delivered to the animal via the digestive tract. Examples of enzymes that can be delivered or incorporated into the compositions of the invention, include, for example, feed enhancing enzymes selected from the group consisting of I-galactosidases, ϑ-galactosidases, in particular lactases, phytases, ϑ-glucanases, in particular endo-ϑ-1,4-glucanases and endo-ϑ-1,3(4)-glucanases, cellulases, xylosidases, galactanases, in particular arabinogalactan endo-1,4-ϑ-galactosidases and arabinogalactan endo-1,3-ϑ-galactosidases, endoglucanases, in particular endo-1,2-ϑ-glucanase, endo-1,3-I-glucanase, and endo-1,3-ϑ-glucanase, pectin degrading enzymes, in particular pectinases, pectinesterases, pectin lyases, polygalacturonases, arabinanases, rhamnogalacturonases, rhamnogalacturonan acetyl esterases, rhamnogalacturonan-I-rhamnosidase, pectate lyases, and I-galacturonisidases, mannanases, ϑ-mannosidases, mannan acetyl esterases, xylan acetyl esterases, proteases, xylanases, arabinoxylanases and lipolytic enzymes such as lipases, phospholipases and cutinases. Phytases in addition to the phytases having an amino acid sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14 can be used in the methods and compositions of the invention.

In a preferred embodiment, the enzyme used in the compositions (*e.g.*, a dietary aid) of the present invention is a phytase enzyme which is stable to heat and is heat resistant and catalyzes the enzymatic hydrolysis of phytate, *i.e.*, the enzyme is able to

renature and regain activity after a brief (*i.e.*, 5 to 30 seconds), or longer period, for example, minutes or hours, exposure to temperatures of above 50 °C.

A "feed" and a "food," respectively, means any natural or artificial diet, meal or the like or components of such meals intended or suitable for being eaten, taken in, digested, by an animal and a human being, respectively. "Dietary Aid," as used herein, denotes, for example, a composition containing agents that provide a therapeutic or digestive agent to an animal or organism. A "dietary aid," typically is not a source of caloric intake for an organism, in other words, a dietary aid typically is not a source of energy for the organism, but rather is a composition which is taken in addition to typical "feed" or "food".

In various aspects of the invention, feed composition are provided that comprise a substantially purified phytase protein having at least thirty contiguous amino acids of a protein having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14; and a phytate-containing foodstuff.. As will be known to those skilled in the art, such compositions may be prepared in a number of ways, including but not limited to, in pellet form with or without polymer coated additives, in granulate form, and by spray drying. By way of non-limiting example, teachings in the art directed to the preparation of feed include International Publication Nos. WO0070034 A1, WO0100042 A1, WO0104279 A1, WO0125411 A1, WO0125412 A1, and EP 1073342A.

An agent or enzyme (*e.g.*, a phytase) may exert its effect *in vitro* or *in vivo*, *i.e.* before intake or in the stomach or gizzard of the organism, respectively. Also a combined action is possible.

Although any enzyme may be incorporated into a dietary aid, reference is made herein to phytase as an exemplification of the methods and compositions of the invention. A dietary aid of the invention includes an enzyme (*e.g.*, a phytase). Generally, a dietary aid containing a phytase composition is liquid or dry.

Liquid compositions need not contain anything more than the enzyme (*e.g.* a phytase), preferably in a highly purified form. Usually, however, a stabilizer such as glycerol, sorbitol or mono propylen glycol is also added. The liquid composition may also comprise other additives, such as salts, sugars, preservatives, pH-adjusting agents, proteins, phytate (a phytase substrate). Typical liquid compositions are aqueous or oil-based slurries. The liquid compositions can be added to a biocompatible composition for slow release. Preferably the enzyme is added to a dietary aid composition that is a biocompatible material (*e.g.*, biodegradable or non-biodegradable) and includes the addition of recombinant cells into, for example, porous microbeads.

Dry compositions may be spray dried compositions, in which case the composition need not contain anything more than the enzyme in a dry form. Usually, however, dry compositions are so-called granulates which may readily be mixed with a food or feed components, or more preferably, form a component of a pre-mix. The particle size of the enzyme granulates preferably is compatible with that of the other components of the mixture. This provides a safe and convenient means of incorporating enzymes into animal feed. Preferably the granulates are biocompatible and more preferably they biocompatible granulates are non-biodegradable.

Agglomeration granulates coated by an enzyme can be prepared using agglomeration technique in a high shear mixer Absorption granulates are prepared by having cores of a carrier material to absorp/be coated by the enzyme. Preferably the carrier material is a biocompatible non-biodegradable material that simulates the role of stones or grit in the gizzard of an animal. Typical filler materials used in agglomeration techniques include salts, such as disodium sulphate. Other fillers are kaolin, talc, magnesium aluminium silicate and cellulose fibres. Optionally, binders such as dextrins are also included in agglomeration granulates. The carrier materials can be any biocompatible material including biodegradable and non-biodegradable materials (*e.g.*, rocks, stones, ceramics, various polymers). Optionally, the granulates are coated with a coating mixture. Such mixture comprises coating agents, preferably hydrophobic coating agents, such as hydrogenated palm oil and beef tallow, and if desired other additives, such as calcium carbonate or kaolin.

Additionally, the dietary aid compositions (*e.g.*, phytase dietary aid compositions) may contain other substituents such as colouring agents, aroma compounds, stabilizers, vitamins, minerals, other feed or food enhancing enzymes etc. A typical additive usually comprises one or more compounds such as vitamins, minerals or feed enhancing enzymes and suitable carriers and/or excipients.

In a one embodiment, the dietary aid compositions of the invention additionally comprises an effective amount of one or more feed enhancing enzymes, in particular feed enhancing enzymes selected from the group consisting of I-galactosidases, ϑ-galactosidases, in particular lactases, other phytases, ϑ-glucanases, in particular endo-ϑ-1,4-glucanases and endo-ϑ-1,3(4)-glucanases, cellulases, xylosidases, galactanases, in particular arabinogalactan endo-1,4-ϑ-galactosidases and arabinogalactan endo-1,3-ϑ-galactosidases, endoglucanases, in particular endo-1,2-ϑ-glucanase, endo-1,3-I-glucanase, and endo-1,3-ϑ-glucanase, pectin degrading enzymes, in particular pectinases, pectinesterases, pectin lyases, polygalacturonases, arabinanases, rhamnogalacturonases, rhamnogalacturonan acetyl esterases, rhamnogalacturonan-I-rhamnosidase, pectate lyases, and I-galacturonisidases, mannanases, ϑ-mannosidases, mannan acetyl esterases, xylan acetyl esterases, proteases, xylanases, arabinoxylanases and lipolytic enzymes such as lipases, phospholipases and cutinases.

The animal dietary aid of the invention is supplemented to the mono-gastric animal before or simultaneously with the diet. In one embodiment, the dietary aid of the invention is supplemented to the mono-gastric animal simultaneously with the diet. In another embodiment, the dietary aid is added to the diet in the form of a granulate or a stabilized liquid.

An effective amount of an enzyme in a dietary aid of the invention is from about 10-20,000; from about 10 to 15,000, from about 10 to 10,000, from about 100 to 5,000, or from about 100 to about 2,000 FYT/kg dietary aid.

Non-limiting examples of other specific uses of the phytase of the invention is in soy processing and in the manufacture of inositol or derivatives thereof.

The invention also relates to a method for reducing phytate levels in animal manure, wherein the animal is fed a dietary aid containing an effective amount of the phytase of the invention. As stated in the beginning of the present application one important effect thereof is to reduce the phosphate pollution of the environment.

In another embodiment, the dietary aid is a magnetic carrier. For example, a magnetic carrier containing an enzyme (*e.g.*, a phytase) distributed in, on or through a magnetic carrier (*e.g.*, a porous magnetic bead), can be distributed over an area high in phytate and collected by magnets after a period of time. Such distribution and recollection of beads reduces additional pollution and allows for reuse of the beads. In addition, use of such magnetic beads *in vivo* allows for the localization of the dietary aid to a point in the digestive tract where, for example, phytase activity can be carried out. For example, a dietary aid of the invention containing digestive enzymes (*e.g.*, a phytase) can be localized to the gizzard of the animal by juxtapositioning a magnet next to the gizzard of the animal after the animal consumes a dietary aid of magnetic carriers. The magnet can be removed after a period of time allowing the dietary aid to pass through the digestive tract. In addition, the magnetic carriers are suitable for removal from the organism after sacrificing or to aid in collection.

When the dietary aid is a porous particle, such particles are typically impregnated by a substance with which it is desired to release slowly to form a slow release particle. Such slow release particles may be prepared not only by impregnating the porous particles with the substance it is desired to release, but also by first dissolving the desired substance in the first dispersion phase. In this case, slow release particles prepared by the method in which the substance to be released is first dissolved in the first dispersion phase are also within the scope and spirit of the invention. The porous hollow particles may, for example, be impregnated by a slow release substance such as a medicine, agricultural chemical or enzyme. In particular, when porous hollow particles impregnated by an enzyme are made of a biodegradable polymers, the particles themselves may be used as an agricultural chemical or fertilizer, and they have no adverse effect on the environment. In one embodiment the porous particles are magnetic in nature.

The porous hollow particles may be used as a bioreactor support, in particular an enzyme support. Therefore, it is advantageous to prepare the dietary aid utilizing a method of a slow release, for instance by encapsulating the enzyme of agent in a microvesicle, such as a liposome, from which the dose is released over the course of several days, preferably between about 3 to 20 days. Alternatively, the agent (*e.g.*, an enzyme) can be formulated for slow release, such as incorporation into a slow release polymer from which the dosage of agent (*e.g.*, enzyme) is slowly released over the course of several days, for example from 2 to 30 days and can range up to the life of the animal.

As is known in the art, liposomes are generally derived from phospholipids or other lipid substances. Liposomes are formed by mono- or multilamellar hydrated liquid crystals that are dispersed in an aqueous medium. Any non-toxic, physiologically acceptable and metabolizable lipid capable of forming liposomes can be used. The present compositions in liposome form can contain stabilizers, preservatives, excipients, and the like in addition to the agent. Some preferred lipids are the phospholipids and the phosphatidyl cholines (lecithins), both natural and synthetic. Methods to form liposomes are known in the art. See, for example, Prescott, Ed., Methods in Cell Biology, Volume XIV, Academic Press, New York, N.Y. (1976), p. 33 et seq.

Also within the scope of the invention is the use of a phytase of the invention during the preparation of food or feed preparations or additives, *i.e.*, the phytase excerts its phytase activity during the manufacture only and is not active in the final food or feed product. This aspect is relevant for instance in dough making and baking. Accordingly, phytase or recombinant yeast expressing phytase can be impregnated in, on or through a magnetic carriers, distributed in the dough or food medium, and retrieved by magnets.

The dietary aid of the invention may be administered alone to animals in an biocompatible (*e.g.*, a biodegradable or non-biodegradable) carrier or in combination with other digestion additive agents. The dietary aid of the invention thereof can be readily administered as a top dressing or by mixing them directly into animal feed or

provided separate from the feed, by separate oral dosage, by injection or by transdermal means or in combination with other growth related edible compounds, the proportions of each of the compounds in the combination being dependent upon the particular organism or problem being addressed and the degree of response desired. It should be understood that the specific dietary dosage administered in any given case will be adjusted in accordance with the specific compounds being administered, the problem to be treated, the condition of the subject and the other relevant facts that may modify the activity of the effective ingredient or the response of the subject, as is well known by those skilled in the art. In general, either a single daily dose or divided daily dosages may be employed, as is well known in the art.

If administered separately from the animal feed, forms of the dietary aid can be prepared by combining them with non-toxic pharmaceutically acceptable edible carriers to make either immediate release or slow release formulations, as is well known in the art. Such edible carriers may be either solid or liquid such as, for example, corn starch, lactose, sucrose, soy flakes, peanut oil, olive oil, sesame oil and propylene glycol. If a solid carrier is used the dosage form of the compounds may be tablets, capsules, powders, troches or lozenges or top dressing as micro-dispersable forms. If a liquid carrier is used, soft gelatin capsules, or syrup or liquid suspensions, emulsions or solutions may be the dosage form. The dosage forms may also contain adjuvants, such as preserving, stabilizing, wetting or emulsifying agents, solution promoters, etc. They may also contain other therapeutically valuable substances.

Thus, a significant advantages of the invention include for example, 1) ease of manufacture of the active ingredient loaded biocompatible compositions; 2) versatility as it relates to the class of polymers and/or active ingredients which may be utilized; 3) higher yields and loading efficiencies; and 4) the provision of sustained release formulations that release active, intact active agents *in vivo*, thus providing for controlled release of an active agent over an extended period of time. In addition, another advantage is due to the local delivery of the agent with in the digestive tract (*e.g.*, the gizzard) of the organism. As used herein the phrase "contained within" denotes a method for formulating an agent into a composition useful for controlled release, over an extended period of time of the agent.

In the sustained-release or slow release compositions of the invention, an effective amount of an agent (*e.g.*, an enzyme or antibiotic) will be utilized. As used herein, sustained release or slow release refers to the gradual release of an agent from a biocompatible material, over an extended period of time. The sustained release can be continuous or discontinuous, linear or non-linear, and this can be accomplished using one or more biodegradable or non-biodegradable compositions, drug loadings, selection of excipients, or other modifications. However, it is to be recognized that it may be desirable to provide for a "fast" release composition, that provides for rapid release once consumed by the organism. It is also to be understood that by "release" does not necessarily mean that the agent is released from the biocompatible carrier. Rather in one embodiment, the slow release encompasses slow activation or continual activation of an agent present on the biocompatible composition. For example, a phytase need not be released from the biocompatible composition to be effective. In this embodiment, the phytase is immobilized on the biocompatible composition.

The animal feed may be any protein-containing organic meal normally employed to meet the dietary requirements of animals. Many of such protein-containing meals are typically primarily composed of corn, soybean meal or a corn/soybean meal mix. For example, typical commercially available products fed to fowl include Egg Maker Complete, a poultry feed product of Land O'Lakes AG Services, as well as Country Game and Turkey Grower a product of Agwa, Inc. (see also The Emu Farmer's Handbook by Phillip Minnaar and Maria Minnaar). Both of these commercially available products are typical examples of animal feeds with which the present dietary aid and/or the enzyme phytase may be incorporated to reduce or eliminate the amount of supplemental phosphorus, zinc, manganese and iron intake required in such compositions.

The present invention is applicable to the diet of numerous animals, which herein is defined as including mammals (including humans), fowl and fish. In particular, the diet may be employed with commercially significant mammals such as pigs, cattle, sheep, goats, laboratory rodents (rats, mice, hamsters and gerbils), fur-bearing animals such as mink and fox, and zoo animals such as monkeys and apes, as

well as domestic mammals such as cats and dogs. Typical commercially significant avian species include chickens, turkeys, ducks, geese, pheasants, emu, ostrich, loons, kiwi, doves, parrots, cockatiel, cockatoo, canaries, penguins, flamingoes, and quail. Commercially farmed fish such as trout would also benefit from the dietary aids disclosed herein. Other fish that can benefit include, for example, fish (especially in an aquarium or acquaculture environment, *e.g.*, tropical fish), goldfish and other ornamental carp, catfish, trout, salmon, shark, ray, flounder, sole, tilapia, medaka, guppy, molly, platyfish, swordtail, zebrafish, and loach.

The following examples are intended to illustrate, but not to limit, the invention. While the procedures described in the examples are typical of those that can be used to carry out certain aspects of the invention, other procedures known to those skilled in the art can also be used.

EXAMPLES

Example 1: Identification and Isolation of Nucleic Acids of the Invention

SEQ ID NO:1 was identified in a Blast search performed using the *E. coli* appa gene as a probe against a plurality of unfinished microbial genomes deposited with GenBank (as described above). A number of hits were identified including a gene found in *Yersinia pestis*, the organism responbsible for bubonic plague.

Standard techniques may be utilized to produce the nucleic acid molecule of SEQ ID NO:3. For example, the appropriate oligonucleotides covering the entire legth of the gene sequence may be synthesized *in vitro* and ligated together. Table 3 presents such a list of oligonucleotides for the construction of a nucleic acid encoding the Y. pestis phytase.

Table 3: Oligonucleotides for the Construction of Y. pestis Phytase

Y2F1F	CTTCTACTAGAATTCAT TAAAGAGGAGAAATTAA CCATGTCCGTACTGGA GAA (SEQ ID NO:19)	**Y2F1R**	ACGCGGTTCTCCAGTA CGGACATGGTTAATTTC TCCTCTTTAATGAATTC TAGTAGAAG (SEQ ID NO:44)
Y2F2F	CCGCGTCCGCCTTTCC GGTTTAGTGTTAATGCT GTCCGGCCTGGCTGC	**Y2F2R**	GGTGATAGCAGCCAGG CCGGACAGCATTAACA CTAAACCGGAAAGGCG

	(SEQ ID NO:20)		G (SEQ ID NO:45)
Y2F3F	TATCACCGCGCCTGTG GCCGCCGAACCATCGG GGTACACCTTAGAACG TGTAG (SEQ ID NO:21)	**Y2F3R**	AAAATAACTACACGTTC TAAGGTGTACCCCGAT GGTTCGGCGGCCACAG GCGC (SEQ ID NO:46)
Y2F4F	TTATTTTGAGTCGCCAT GGTGTGCGTAGCCCGA CTAAGCAGACGCAGCT GATGAA (SEQ ID NO:22)	**Y2F4R**	ACATCATTCATCAGCTG CGTCTGCTTAGTCGGG CTACGCACACCATGGC GACTC (SEQ ID NO:47)
Y2F5F	TGATGTAACACCTGATA AGTGGCCTCAGTGGCC GGTTAAAGCGGGCTAT TTGACTCCTCGTGGC (SEQ ID NO:23)	**Y2F5R**	CGGCGCCACGAGGAGT CAAATAGCCCGCTTTAA CCGGCCACTGAGGCCA CTTATCAGGTGTT (SEQ ID NO:48)
Y2F6F	GCCGAACTGGTCACCC TGATGGGCGGGTTCTA TGGCGATTATTTCCGCA GTTTGGGTCTTTTGGC CG (SEQ ID NO:24)	**Y2F6R**	CGGCCAAAAGACCCAA ACTGCGGAAATAATCG CCATAGAACCCGCCCA TCAGGGTGACCAGTT (SEQ ID NO:49)
Y2F7F	GCCGCGGGCTGCCCG GCAGAGGGCGGTGTAT ATGCACAGGCAGATAT CGACCAGCG (SEQ ID NO:25)	**Y2F7R**	CGAGTGCGCTGGTCGA TATCTGCCTGTGCATAT ACACCGCCCTCTGCCG GGCAGCCCGCGGC (SEQ ID NO:50)
Y2F8F	CACTCGTTTAACCGGT CAGGCTTTTCTGGATG GTGTGGCGCCGGGTTG CGGCCTG (SEQ ID NO:26)	**Y2F8R**	GACAGTCAGGCCGCAA CCCGGCGCCACACCAT CCAGAAAAGCCTGACC GGTTAAA (SEQ ID NO:51)
Y2F9F	ACTGTCCACAATCAGG CCGATCTTAAGAAAACC GATCCTCTGTTTCATCC (SEQ ID NO:27)	**Y2F9R**	TCAACGGGATGAAACA GAGGATCGGTTTTCTTA AGATCGGCCTGATTGT G (SEQ ID NO:52)
Y2F10F	CGTTGAAACCGGCGTC TGTAAACTGGACAACG CCCAAACCGATAAGGC AATTGA (SEQ ID NO:28)	**Y2F10R**	GCGTTCCTCAATTGCCT TATCGGTTTGGGCGTT GTCCAGTTTACAGACG CCGGTT (SEQ ID NO:53)
Y2F11F	GGAACGCCTGGGCGG CCCGTTAGACACGGTA AGCCAGCGCTATGCCA AACCGTTTGCG (SEQ ID NO:29)	**Y2F11R**	CCATTTGCGCAAACGG TTTGGCATAGCGCTGG CTTACCGTGTCTAACG GGCCGCCCAG (SEQ ID NO:54)
Y2F12F	CAAATGGGCGATGTCC TGAACTTCGCTGCGAG TCCGTACTGCAAGTCA CTGCAGCAGCAGGGGA AAA (SEQ ID NO:30)	**Y2F12R**	TTTTCCCCTGCTGCTGC AGTGACTTGCAGTACG GACTCGCAGCGAAGTT CAGGACATCGC (SEQ ID NO:55)
Y2F13F	AAAACTTGTGACTTCGC ACACTTTGCGGCCAAC	**Y2F13R**	TCGTGCCTTCCTTGTTT ACATTAACTTCGTTGGC

	GAAGTTAATGTAAACAA GGAAG (SEQ ID NO:31)		CGCAAAGTGTGCGAAG TCACAAGTTTT (SEQ ID NO:56)
Y2F14F	GCACGAAAGTTACCCT GTCAGGCCCCCTGGCG CTGTCTAGCACGTTGG GCGAAATCTT (SEQ ID NO:32)	**Y2F14R**	AGCAAGAAGATTTCGC CCAACGTGCTAGACAG CGCCAGGGGGCCTGAC AGGGTAACTT (SEQ ID NO:57)
Y2F15F	CTTGCTGCAGAACGCG CAGGCGATGCCCGAAG TAGCGTGGCAGCGTTT GAAAGGCGCT (SEQ ID NO:33)	**Y2F15R**	GTTCTCAGCGCCTTTCA AACGCTGCCACGCTAC TTCGGGCATCGCCTGC GCGTTCTGC (SEQ ID NO:58)
Y2F16F	GAGAACTGGGTGTCTC TTCTGAGCCTGCACAAT GCACAGTTCAACCTGA TGGCTAAAA (SEQ ID NO:34)	**Y2F16R**	ATGGCGTTTTAGCCATC AGGTTGAACTGTGCATT GTGCAGGCTCAGAAGA GACACCCA (SEQ ID NO:59)
Y2F17F	CGCCATACATTGCACG CCACAAAGGCACGCCG CTTTTACAGCAAATCGA TACCGCACTGA (SEQ ID NO:35)	**Y2F17R**	GCAGGGTCAGTGCGGT ATCGATTTGCTGTAAAA GCGGCGTGCCTTTGTG GCGTGCAATGT (SEQ ID NO:60)
Y2F18F	CCCTGCAACTGGACGC CCAGGGGCAAAAACTG CCGATCTCGGCTCAGA ACCGTGTTTTATTCCTG GGTG (SEQ ID NO:36)	**Y2F18R**	CACCCAGGAATAAAAC ACGGTTCTGAGCCGAG ATCGGCAGTTTTTGCC CCTGGGCGTCCAGTT (SEQ ID NO:61)
Y2F19F	GGTGGCCACGACACAA ATATTGCTAACATCGCC GGTATGCTGGGCGCAG ATTGGCAGTTAC (SEQ ID NO:37)	**Y2F19R**	GTTCCGGTAACTGCCA ATCTGCGCCCAGCATA CCGGCGATGTTAGCAA TATTTGTGTCGTGGCCA CC (SEQ ID NO:62)
Y2F20F	CGGAACAACCGGATAA CACCCCACCGGGCGGC GGTCTGGTCTTTGAGC TGTGGCAGAAT (SEQ ID NO:38)	**Y2F20R**	GTCCGGATTCTGCCAC AGCTCAAAGACCAGAC CGCCGCCCGGTGGGG TGTTATCCGGTT (SEQ ID NO:63)
Y2F21F	CCGGACAATCATCAAC GTTATGTGGCCGTTAA GATGTTCTATCAGACCA TGGAT (SEQ ID NO:39)	**Y2F21R**	GCAGTTGATCCATGGT CTGATAGAACATCTTAA CGGCCACATAACGTTG ATGATT (SEQ ID NO:64)
Y2F22F	CAACTGCGTAACGCCG AGAAGCTGGATTTAAA GAACAATCCCGCCGGC ATCATCAGTG (SEQ ID NO:40)	**Y2F22R**	CAGCGACACTGATGAT GCCGGCGGGATTGTTC TTTAAATCCAGCTTCTC GGCGTTAC (SEQ ID NO:65)
Y2F23F	TCGCTGTGGCCGGCTG CGAGAATAATGGTGAC GATAAACTGTGCGAAC	**Y2F23R**	AAGTATCAAGTTCGCAC AGTTTATCGTCACCATT ATTCTCGCAGCCGGCC

	TTG (SEQ ID NO:41)		A (SEQ ID NO:66)
Y2F24F	ATACTTTTCAAAAAAAA GTAGCGAAAGTCATTG AACCTGCGTGTCATATT AAAGCTTCTACTACTA (SEQ ID NO:42)	**Y2F24R**	TAGTAGTAGAAGCTTTA ATATGACACGCAGGTT CAATGACTTTCGCTACT TTTTTTTGAA (SEQ ID NO:67)

Briefly, the *Yersinia pestis* synthetic phytase gene sequence was produced by first synthesizing all fragments provided for by each forward and reverse oligo pair presented in Table 3. The reaction conditions for the synthesis of these fragments was as follows. The gel purified primers (IDT, 200 nmole synthesis, polyacrylamide gel electrophoresis (PAGE) purified) were resuspended in H_2O at 100 pMoles/ul, and equal amounts of forward and reverse primers were mixed together. The primers were annealed in a thermocycler under the following conditions: 5 min 94°C; 5 min 72°C; 5min 60°C; 5 min 50°C; 5 min 37°C; and 5 min 16°C. Equal amounts of homologous fragments were mixed after checking the concentration of each. First, the samples were diluted from a concentration of 50 pMoles/ul (all genes) to 5 pMoles/µl, and load 2 µl of each sample were loaded onto an agarose gel to check that the relative concentrations were all the same.

The fragments were then subjected to ligation in order to assemble the full length gene. 24 pairs of forward and reverse oligos were used to construct the full length gene.

The full length gene was the ligation product of 4 fragments each consisting of 6 annealed oligo pairs. Once the forward and reverse oligos are annealed they form a double stranded piece of DNA with a compatible overhang for ligation to the next oligo pair.

Gene assembly followed the following protocol: Fragment 1 = ligation product of oligos Y2f1-Y2f6; Fragment 2 = ligation product of oligos Y2f7- Y2f12; Fragment 3 = ligation product of oligos Y2f13 – Y2f18; Fragment 4 = ligation product of oligos Y2f19 – Y2f24.

The ligation reaction consisted of: (1) DNA fragments in 60ul (10ul each of 6 annealed oligos), (2) BRL 5X ligation buffer (16ul), and BRL T4 ligase (1u/ul) 4ul). Samples were incubated for more than 1 hour at 22C.

The full length product was isolated on a 4% agarose gel. Because these are the ligation products of 6 oligos which are each ~50bp, the final product should be ~300-350bp. Next, PCR was used to amplify each fragment with end primers to make more usable material. Each primer had a restriction site designed in to created a compatible overhang for subsequent ligation to one another.

PCR amplification was done according to the following. For Fragment 1, the following primers were used Y2f ecoR1 (CTA CTA GAA TTC ATT AAA GAG GAG) (SEQ ID NO:68) and Y2BsmB1-6r (TAC TGA CGT CTC ACG GCC AAA AGA CCC AAA CTG CG) (SEQ ID NO:69). Fragment 2 was amplified with primers Y2BsmB1-7f (TAC TGA CGT CTC AGC CGC GGG CTG CCC GGC AGA GG) (SEQ ID NO:70) and Y2BsmB1-12r (TAC TGA CGT CTC ATT TTC CCC TGC TGC TGC AGT GA) (SEQ ID NO:71). Fragment 3 was amplified with primers (Y2BsmB1-13f (TAC TGA CGT CTC AAA AAC TTG TGA CTT CGC ACA CT) (SEQ ID NO:72) and Y2BsmB1-18r (TAC TGA CGT CTC ACA CCC AGG AAT AAA ACA CGG TT) (SEQ ID NO:73). Fragment 4 was amplified with primers Y2BsmB1-19f (TAC TGA CGT CTC AGG TGG CCA CGA CAC AAA TAT TG) (SEQ ID NO:74) and Y2RhinD3 (AGT AGT AGA AGC TTA AAT ATG AC) (SEQ ID NO:75).

The following condition were utilized for PCR reactions:

Template	1 µl of ligation
Forward Primer	40 pMoles
Reverse Primer	40 pMoles
DNTPs	1 µl of 20 mM/dntp Mix (Pharmacia)
PFu polymerase	1 µl of 2.5 U/µl (Stratagene)
10x Pfu buffer	10 µl
Water	X µl to bring up final reaction to 100 µl.

The PCR program was as follows: 95°C for 20 sec; 50°C for 1 min; 72°C for 1 min for a total of 30 cycles.

After amplification of fragments 1-4, digest with appropriate restriction sites and gel isolated. The gene was assembled by first ligating fragment 1 to fragment 2 and fragment 3 to fragment 4. The following conditions were used for the ligation reaction:

Frag1/3	50 µl

Frag2/4	50 µl
5X BRL lig buffer	30 µl
BRL T4 Ligase	20 µl

The sample was incubated at room temperature for more than 1 hour. A sample of this ligation, 1ul, was used as template for another PCR amplification.

Fragment 1+2 reaction were amplified with primers Y2fecoR1 + Y2BsmB1-12R (sequences above), and the fragment 3+4 reaction was amplified with primers Y2BsmB1-13f + Y2BsmB1-24R (sequences above). The PCR products were digested with the appropriate enzymes and gel isolated.

Final assembly was achieved by ligating fragment 1+2 to fragment 3+4 to create the full length gene. This was accomplished utilizing the same ligation reaction conditions as previous previously described. A sample of this ligation reaction,1ul, was amplified with primers Y2fecoR1 and Y2RhinD3. The resulting fragment was digested with EcoR1 and HindIII restriction enzymes. The sample was then gel isolated and ligated into pQE60 (also cut with EcoR1 and HindIII). A sample of this ligation reaction,2ul, was used to transform 40ul of phy635 electrocomp cells. Transformants were then screened for phytase activity. One nucleic acid clone (SEQ ID NO:11) was found, which encoded a protein having the amino acid sequence of SEQ ID NO:12.

Alternatively, the sequence can also be obtained by PCR amplification from *Yersinia pestis* DNA. The selection of appropriate primers and reaction conditions for such an amplification are well within the skill of those in the art.

The original phytase sequence from the unfinished *Yersinia pestis* genome was incomplete for several amino acids. These amino acids occurred at positions 157, 163, 164, and 174 of SEQ ID NO:2. These residues were changed when a synthetic gene (SEQ ID NO:3) was made that included the corresponding amino acids of the *E. coli* appa phytase substituted in place of those residues missing from the *Yersinia pestis* sequence. These changes are identified in bold in Figure 5.

Additional novel phytase gene sequences were identified through library screening. Clone 953-6 (SEQ ID NO:5) and clone 954-2 (SEQ ID NO:9) were isolated from novel, mixed bacterial population libraries constructed from environmental samples (see U.S. Patent No. XXXXXX). In addition, a Rhizobium phytase gene (SEQ ID NO:7) was isolated from a Rhizobium gene library.

Utilizing the sequences disclosed herein, the novel phytase-encoding nucleic acid molecules of the invention can be obtained by a variety of methods known to one skilled in the art. For example, primers can be selected from the Rhizobium sequence provided herein and utilized for the direct PCR amplication of these sequences from genomic DNA. Alternatively, SEQ ID NOS:1, 3, 5, 7, and 9 can be produced synthetically through ligation of artificial oligonucleotides that span the entire length of these sequences.

Example 2: Recombinant Expression of Phytase Proteins

In order to express the isolated phytase proteins of the invention in yeast and Psuedomonas, the nucleic acid expression vectors must first be introduced into the desired host.

Plasmid DNA Transformation Protocol for *Pseudomonas*

Electroporation competent Psuedomonas cells were prepared according to the following protocol. One milliliter of an overnight culture was innoculated into 100 ml LB, and the culture was incubated in a 30°C shaker flask until an OD 600 reading of 0.5-0.7. Next, the bacteria are harvested by spinning at 3000 rpm for 10 minutes at 4°C. The resulting cell pellet was washed with 100 ml ice-cold dd$H_2$0 and spun at 3000 rpm for 10 minutes at 4°C to collect the cells. The washing was repeated. The cells were then washed with 50 ml 10% ice-cold glycerol(in dd$H_2$0) once and collected by spinning at 3000 rpm for 10 minutes at 4°C. The bacterial cell pellet was resuspended into 2 ml ice-cold 10% glycerol(in dd$H_2$0) The cells were aliquoted (50 µl or 100 µl) into tubes and stored at -80°C.

Electroporated was done with 1ul plasmid DNA mixed with 50 µl competent cell and kept on ice for 5 minutes. The mixture was transferred to a pre-chilled cuvette(0.2 cm gap, Bio-Rad). The DNA was transformed into bacteria by

electroporation with Bio-Rad machine. (Setting: Volts: 2.25KV; time: 5ms; capacitance: 25 μF)

300 μl SOC medium is added to the cell mixture and bacteria are incubated at 30°C in a shaker flask for one hour. A certain amount of culture is spread on LA plate with antibiotics and the plates were incubated at 30°C.

Plasmid DNA Transformation Protocol for Yeast

One day before the experiment, 10 ml of YPD medium was inoculated with a single yeast colony of the strain to be transformed. It was grown overnight to saturation at 30°C. On the day of competent cell preparation, the total volume of yeast overnight culture was transferred to a 2 L baffled flask containing 500 ml YPD medium. The culture was grown with vigorous shaking at 30°C to an $OD_{600} \cong$ 0.8-1.0.

500 ml of culture was harvested by centrifuging at 4000 x g, 4°C, for 5 min in autoclaved bottles. The supernatant was subsequently dwascarded. The cell pellet was washed in 250 ml cold sterile water. Washing was repeated twice. The supernatant was dwascarded.

The pellet was resuspended in 30 ml of ice-cold 1M Sorbitol. The suspension was transferred into a sterile 50 ml conical tube. The mixture was centrifuged in a GP-8 centrifuge 2000 rpm, 4°C for 10 min. The supernatant was dwascarded.

The pellet was resuspended in 50μl of ice-cold 1M Sorbitol. The final volume of resuspended yeast should be 1.0 to 1.5 ml and the final OD600 should be ~200.

In a sterile, ice-cold 1.5-ml microcentrifuge tube, 40ul concentrated yeast cells were mixed with 1ug of DNA contained in ≤5 μl. The mixture was transferred to an ice-cold 0.2-cm-gap disposable electroporation cuvette and pulsed at 1.5 kV, 25 uF, 200Ω. It should be noted that the time constant reported by the Gene Pulser will vary from 4.2 to 4.9 msec. Times <4 msec or the presence of a current arc (evidenced by a spark and smoke) indicate that the conductance of the yeast/DNA mixture was too high.

400 µl ice-cold 1M sorbitol was added to the cuvette and the yeast was recovered, with gentle mixing. 200 µl aliquots of the east suspension should be spread directly on sorbitol selection plates. Incubate 3 to 6 days at 30°C until colonies appear.

The sytnthetic gene was assayed using both a micortitre based molybdate asay described herein or a plate based screen using a phytate overlay (Golovan et al. (2000) *Can. J. Microbiol.* 46: 59-71).

Figure X prsents results of an experiment designed to construct a synthetic codon-optimized *Y. pestis* phytase gene. The gene sequence construct as described herein was subsequently ligated into the pQE60 expression plasmid vector and transformed into PHY635 host cells. Colonies from this ligation were assayed with the phytate overlay method to screen for phytase activity.

A phytate-clearing colony was identified. This colony was cored from the agar and plasmid DNA was isolated and used to transform two hosts: TOP10 and TOP10F'. Figure X presents results of a phytase overlay screen on these cell types transformed with the synthetic Y. pestis phytase encoding nucleic acid. Isolates 1-10 were from the transformation performed in TOP10 host; and isolates 11-20 were from the transformation performed in TOP10F' host. Vector control is shown in the lower right corner (pQE60). These results demonstrate that clones with phytase activity result in clearing of the pytate overlay.

The above are additional isolates from the re-transformation described in Figure 1. Ed1#21 OL is in the TOP10 host; Ed1#22 OL (SEQ ID NO:11) is in the TOP10F' host. This figure shows that the clone expressing SEQ ID NO:11 displays phytase activty. As a result, the clone carrying SEQ ID NO:11 was selected and the insert was then sequenced.

Example 3: Glycosylation Stabilizes Phytase

Experiments were conducted to evaluate the affect of glycosilation on the half life of phytase enzyme exposed to pepsin, a gastrointestinal enzyme. Studies were first

undertaken to determine the type of glycosylation on phytase expressed in pichia and yeast.

To remove O-glycosylated chains, 1 mU of O-glycosidase (Roche Molecular Biochemicals, Germany) was added to 50 µg of phytase in a buffer containing 20 mM Tris, pH 7.5 followed by incubation at 37 °C overnight. To remove N-glycosilated chains, 50 mU of Endoglycosidase H (Roche Molecular Biochemicals, Germany) was added to 50 µg of phytase in a buffer containing 50 mM sodium phosphate, pH 6.5 and incubated at 37 °C overnight. After digestion, 1 µg of the protein was checked on a 12% Tris-Glycine Gel (Invitrogen, San Diego, CA). The results are presented in Figure

For mass spectral analysis, all proteins need to be denatured, reduced and alkalized. In detail, equal volume of 8 M urea (Sigma, MI) was added to phytase solution and incubated at 37 °C for 30 min. To reduce the protein, freshly made DTT (10 mg/mL) (Sigma, MI) was added to this mixture at a final concentration of 0.04 mg/mL followed by an incubation at 37 °C for 30 minutes. Next, 20 mg/mL of Iodoacetamide (Sigma, MI) was added to the reduced protein mixture at a final concentration of 20 µg/mL and incubated at 37 °C for 30 min for alkylation.

After the phytase protein was denatured, reduced and alkalized, the protein was then dialyzed into a buffer containing 34 mM NaCl and 0.08 N HCl. Pepsin (5-20 mg/mL) was added to digest phytase at 37 °C overnight. The complete digestion of the protein can be analyzed by SDS-PAGE.

Phytase fragments digested by pepsin were loaded on a Con A column (Pharmacia Biotech, Piscataway, NJ) in a buffer containing 20 mM Tris, pH 7.4, 0.5 M NaCl, 1 mM CaCl2, 1 mM MnCl2, and 1 mM MgCl2. The column was washed extensively with the same buffer. The glycosylated peptides were eluted using 20 mM Tris buffer pH 7.5 containing 0.5 M D-Methylmannoside.

For MALDI mass spectral analysis, two types of matrices were used in these experiments for either peptides or protein analysis. 3,5-Dimethoxy-4-hydroxycinnamic acid (10 mg/ml) dissolved in 49.9% water, 50% methanol, and 0.1% TFA was used for protein analysis. Alpha-Cyano-4-hydroxycinnamic acid (10 mg/ml) dissolved in 50%

methanol, 49.9% ethanol and 0.1% TFA was used for peptide analysis. To apply on a steel probe tip, 1 μL of sample was mixed well with 1 μL of matrix solution. The samples mixed with matrix were air dried on the probe and analyzed on a Voyager-DE STR instrument (PE Biosystems, Foster City, CA).

The prediction of glycosylated sites of phytase was done using the Post-translational Modification Prediction program at website www.expasy.ch. The glycosylated peptide identification was mapped by PeptideMass program in the same website.

LITERATURE CITED

(The teachings of all references cited in this application were hereby incorporated by reference in their entirety unless otherwise indicated.)

Association of Official Analytical Chemists: Official Methods of Analysis. Association of Official Analytical Chemists, Washington, DC., 1970.

Ausubel FM, *et al.* Current Protocols in Molecular Biology. Greene Publishing Assoc., Media, PA. ©1987, ©1989, ©1992.

Barnes WM: PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proceedings of the National Academy of Sciences, USA* 91(6):2216-2220, 1994.

Bayer EA, Morag E, Lamed R: The cellulosome--a treasure-trove for biotechnology. *Trends Biotechnol* 12(9):379-86, (Sep) 1994.

Bevan M: Binary Agrobacterium vectors for plant transformation. *Nucleic Acids Research* 12(22):8711-21, 1984.

Bird *et al.* *Plant Mol Biol* 11:651, 1988..

Blobel G, Walter P, Chang CN, Goldman BM, Erickson AH, Lingappa VR: Translocation of proteins across membranes: the signal hypothesis and beyond. *Symp Soc Exp Biol* 33:9-36, 1979.

Brederode FT, Koper-Zawrthoff EC, Bol JF: Complete nucleotide sequence of alfalfa mosaic virus RNA 4. *Nucleic Acids Research* 8(10):2213-23, 1980.

Clark WG, Register JC 3d, Nejidat A, Eichholtz DA, Sanders PR, Fraley RT, Beachy RN: Tissue-specific expression of the TMV coat protein in transgenic tobacco plants affects the level of coat protein-mediated virus protection. *Virology* 179(2):640-7, (December) 1990.

Cole, *et al.*: Monoclonal Antibodies and Cancer Therapy. A.R. Liss, New York. ©1985.

Coligan JE, *et al.*: Current Protocols in Immunology. J. Wiley and Sons, New York. ©1996.

Coruzzi G, Broglie R, Edwards C, Chua NH: Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase. *EMBO J* 3(8):1671-9, 1984.

Cosgrove DJ: Inositol phosphate phosphatases of microbiological origin. Inositol phosphate intermediates in the dephosphorylation of the hexaphosphates of myo-inositol, scyllo-inositol, and D-chiro-inositol by a bacterial (Pseudomonas sp.) phytase. *Aust J Biol Sci* 23(6):1207-1220, 1970.

Dassa E, Cahu M, Desjoyaux-Cherel B, Boquet PL: The acid phosphatase with optimum pH of 2.5 of Escherichia coli. Physiological and Biochemical study. *J Biol Chem* 257(12):6669-76, (Jun 25) 1982.

Davis LG, *et al.* Basic Methods in Molecular Biology. Elsevier, New York, ©1986.

Duarte JC, Costa-Ferreira M: Aspergilli and lignocellulosics: enzymology and biotechnological applications. *FEMS Microbiol Rev* 13(2-3):377-86, (Mar) 1994.

Food Chemicals Codex, 4th Edition. Committee on Food Chemicals Codex, Food and Nutrition Board, Institute of Medicine, National Academy of Sciences. Published: National Academy Press, Washington, DC, ©1996.

Garcia PD, Ghrayeb J, Inouye M, Walter P: Wild type and mutant signal peptides of Escherichia coli outer membrane lipoprotein interact with equal efficiency with mammalian signal recognition particle. *J Biol Chem* 262(20):9463-8, (July 15) 1987.

Gluzman Y: SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* 23(1):175-182, 1981.

Goeddel DV, Shepard HM, Yelverton E, Leung D, Crea R, Sloma A, Pestka S: Synthesis of human fibroblast interferon by E. Coli. *Nucleic Acids Research* 8(18):4057-4074, 1980.

Gordon-Kamm WJ, Spencer TM, Mangano ML, Adams TR, Daines RJ, Start WG, O'Brien JV, Chambers SA, Adams Jr. WR, Willets NG, Rice TB, Mackey CJ, Krueger RW, Kausch AP, Lemaux PG. *Plant Cell* 2:603, 1990..

Graf E: Phytic Acid: Chemistry and Applications. Pilatus Press, Minneapolis. 1986.

Greiner R, Haller E, Konietzny U, Jany KD: Purification and characterization of a phytase from Klebsiella terrigena. *Arch Biochem Biophys* 341(2):201-6, (May 15) 1997.

Greiner R, Konietzny U: Construction of a bioreactor to produce special breakdown products of phytate. *J Biotechnol* 48(1-2):153-9, (July 18) 1996.

Greiner R, Konietzny U, Jany KD: Purification and characterization of two phytases from Escherichia coli. *Arch Biochem Biophys* 303(1):107-13, (May 15) 1993.

Guilley H, Dudley RK, Jonard G, Balazs E, Richards KE: Transcription of Cauliflower mosaic virus DNA: detection of promoter sequences, and characterization of transcripts. *Cell* 30(3):763-73, 1982.

Hespell RB, Whitehead TR: Physiology and genetics of xylan degradation by gastrointestinal tract bacteria. *J Dairy Sci* 73(10):3013-22, (Oct) 1990.

Hoekema A, Hirsch PR, Hooykaas PJJ, Schilperoort RA. *Nature* 303:179, 1983.

Horsch RB, Fry JE, Hoffmann NL, Eichholtz D, Rogers SG, Fraley RT. *Science* 227:1229, 1985.

Igarashi M, Hollander VP: Acid phosphatase from rat liver. Purification, crystallization, and properties. *J Biol Chem* 243(23):6084-9, (Dec. 10) 1968.

International Union of Biochemistry and Molecular Biology, Nomenclature Committee: Enzyme nomenclature 1992: recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes / prepwered for NC-IUBMB by Edwin C. Webb. Academic Press, c1992.

Jeffries TW: Biochemistry and genetics of microbial xylanases. *Curr Opin Biotechnol* 7(3):337-42, (Jun.) 1996.

Klee HJ, Muskopf YM, Gasser CS: Cloning of an Arabidopsis thaliana gene encoding 5-enolpyruvylshikimate-3-phosphate synthase: sequence analysis and manipulation to obtain glyphosate-tolerant plants. *Mol Gen Genet* 210(3):437-42, (Dec) 1987.

Kohler G, Milstein C: Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 256(5517):495-497, 1975.

Koster-Topfer M, Frommer WB, Rocha-Sosa M, Rosahl S, Schell J, Willmitzer L: A class II patatin promoter is under developmental control in both transgenic potato and tobacco plants. *Mol Gen Genet* 219(3):390-6, (Nov) 1989.

Kozbor. *Immunology Today* 4:72, 1983.

Lee B, Murdoch K, Topping J, Kreis M, Jones MG: Transient gene expression in aleurone protoplasts isolated from developing caryopses of barley and wheat. *Plant Mol Biol* 13(1):21-9, 1989.

National Research Council: Nutrient Requirements of Poultry (9th Revised ed.). National Academy Press, Washington, DC, 1994.

Nayini NR, Markakis P: *Lebensmittel Wissenschaft und Technologie* 17:24-26, 1984.

NCBI, National Library of Medicine. National Institutes of Health: BLAST Sequence Similarity Searching (web site = www.ncbi.nlm.nih.gov).

Nelson TS, Shieh TR, Wodzinski RJ, Wwere JH: Effect of supplemental phytase on the utilization of phytate phosphorus by chicks. *J Nutr* 101(10):1289-1293, 1971.

Ng DT, Walter P: Protein translocation across the endoplasmic reticulum. *Curr Opin Cell Biol* 6(4):510-6, (Aug), 1994.

Potrykus I: Gene transfer methods for plants and cell cultures. *Ciba Found Symp* 154:198-208; discussion 208-12, 1990.

Powar VK, Jagannathan V: Purification and properties of phytate-specific phosphatase from Bacillus subtilis. *J Bacteriol* 151(3):1102-1108, 1982.

Powers T, Walter P: The nascent polypeptide-associated complex modulates interactions between the signal recognition particle and the ribosome. *Curr Biol* 6(3):331-8, (March 1), 1996.

Prade RA: Xylanases: from biology to biotechnology. *Biotechnol Genet Eng Rev*;13:101-31, 1996.

Ryan AJ, Royal CL, Hutchinson J, Shaw CH: Genomic sequence of a 12S seed storage protein from oilseed rape (Brassica napus c.v. jet neuf). *Nucl Acids Res* 17(9):3584, 1989.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239(4839):487-491, 1988.

Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, Cold Spring Habor Press, Cold Spring Habor, NY, ©1989.

SAS: Statistics In: SAS User's Guide (1984 ed.). SAS Institute, Cwerey, NC, 1984.

Schoner FJ, Hope PP, Schwarz G, Wiesche H: Comparative effects of microbial phytase and inorganic phosphorus on performance and retention of phosphorus, calcium, and crude ash in broilers. *J Anim Physiol Anim Nutr* 66:248, 1991.

Schoner FJ, Hope PP, Schwarz G, Wiesche H: Effects of microbial phytase and inorganic phosphate in broiler chicken: Performance and mineral retention at various calcium levels. *J Anim Physiol Anim Nutr* 69:235, 1993.

Shieh TR, Wodzinski RJ, Wwere JH: Regulation of the formation of acid phosphatases by inorganic phosphate in Aspergillus ficuum. *J Bacteriol* 100(3):1161-5, (Dec) 1969.

Shimamoto K, Miyazaki C, Hashimoto H, Izawa T, Itoh K, Terada R, Inagaki Y, Iida S: Trans-activation and stable integration of the maize transposable element Ds

cotransfected with the Ac transposase gene in transgenic rice plants. *Mol Gen Genet* 239(3):354-60, (June) 1993.

Shimizu M: *Bioscience, Biotechnology, and Biochemistry* 56:1266-1269, 1992.

Sijmons PC, Dekker BM, Schrammeijer B, Verwoerd TC, van den Elzen PJ, Hoekema A: Production of correctly processed human serum albumin in transgenic plants. *Biotechnology (N Y)* 8(3):217-21, 1990.

Simons PC, Versteegh HA, Jongbloed AW, Kemme PA, Slump P, Bos KD, Wolters MG, Beudeker RF, Verschoor GJ: Improvement of phosphorus availability by microbial phytase in broilers and pigs. *Br J Nutr* 64(2):525-540, 1990.

Smeekens S, Weisbeek P, Robinson C: Protein transport into and within chloroplasts. *Trends Biochem Sci* 15(2):73-6, 1990.

Smith AG, Gasser CS, Budelier KA, Fraley RT: Identification and characterization of stamen- and tapetum-specific genes from tomato. *Mol Gen Genet* 222(1):9-16, (June)1990.

Tague BW, Dickinson CD, Chrispeels MJ: A short domain of the plant vacuolar protein phytohemagglutinin targets invertase to the yeast vacuole. *Plant Cell* 2(6):533-46, (June) 1990.

Tingey SV, Walker EL, Corruzzi GM: Glutamine synthetase genes of pea encode distinct polypeptides which were differentially expressed in leaves, roots and nodules. *EMBO J* 6(1):1-9, 1987.

Ullah AH: Production, rapid purification and catalytic characterization of extracellular phytase from Aspergillus ficuum. *Prep Biochem* 18(4):443-458, 1988.

Ullah AH, Gibson DM: Extracellular phytase (E.C. 3.1.3.8) from Aspergillus ficuum NRRL 3135: purification and characterization. *Prep Biochem* 17(1):63-91, 1987

Van den Broeck G, Timko MP, Kausch AP, Cashmore AR, Van Montagu M, Herrera-Estrella L: Targeting of a foreign protein to chloroplasts by fusion to the transit peptide from the small subunit of ribulose 1,5-bisphosphate carboxylase. *Nature* 313(6001):358-63, 1985.

Vasil IK, Vasil V: Totipotency and embryogenesis in plant cell and tissue cultures. *In Vitro* 8(3):117-27, (Nov-Dec) 1972.

Vasil V, Vasil IK: Regeneration of tobacco and petunia plants from protoplasts and culture of corn protoplasts. *In Vitro* 10:83-96, (Jul-Aug) 1974.

Von Heijne G: Towards a comparative anatomy of N-terminal topogenic protein sequences. *J Mol Biol* 189(1):239-42, 1986.

*Walter P, Blobel G. *Biochem Soc Symp* 47:183, 1986.

Wenzler H, Mignery G, Fisher L, Park W: Sucrose-regulated expression of a chimeric potato tuber gene in leaves of transgenic tobacco plants. *Plant Mol Biol* 13(4):347-54, 1989.

Wolter FP, Fritz CC, Willmitzer L, Schell J, Schreier PH rbcS genes in Solanum tuberosum: conservation of transit peptide and exon shuffling during evolution. *Proc Natl Acad Sci U S A* 85(3):846-50, (Feb.) 1988.

Wong KK, Tan LU, Saddler JN: Multiplicity of beta-1,4-xylanase in microorganisms: functions and applications. *Microbiol Rev* 52(3):305-17, (Sept.) 1988.

Yamada K, *et al.*: *Agricultural and Biological Chemistry* 32:1275-1282, 1968.

U.S. Patent No. 3,297,548; Filed July 28, 1964; Issued Jan. 10, 1967. Wwere JH, Bluff L, Shieh TK: Preparation of acid phytase.

U.S. Patent No. 4,946,778; Filed Jan. 19, 1989; Issued Aug. 7, 1990. Ladner RC, Bird RE, Hardman K: Single polypeptide chain binding molecules.

EP0 120,516; Filed Feb. 21, 1984; Issued Oct. 3, 1984. Schilperoort RA, Hoekema A, Hooykaas RJJ: A process of the incorporation of foreign DNA into the genome of dicotyledonous plants; Agrobacterium tumefaciens bacteria and a process for the production thereof; plants and plant cells with modified genetic properties; a process for the preparation.

EP0 321,004; Filed Oct. 28, 1988; Issued Jan. 22, 1992. Vaara T, Vaara M, Simell M, Lehmussaari A, Caransa A: A process for steeping cereals with a new enzyme preparation.

IPN WO 91/05053; Filed Sept. 27, 1990; Issued April 18, 1991. VanGorcom R, *et al.*: Cloning and expression of microbial phytase.

Plant Cell Culture Protocols (Methods in Molecular Biology (Cloth), 111) by Robert D. Hall (Editor) (March 1999) Humana Press; ISBN: 0896035492

Plant Molecular Biology: Essential Techniques by P. Jones (Editor), J. M. Sutton (Editor), Mark Sutton (Contributor) (September 25, 1997) John Wiley & Son Ltd; ISBN: 0471972681

Plant Biochemistry and Molecular Biology by Hans-Walter Heldt (April 1998) Oxford University Press; ISBN: 019850179X

Biochemistry and Molecular Biology of Plants by Bob B. Buchanan (Editor), Wilhelm Gruissem (Editor), Russell L. Jones (July 2000) Amer Society of Plant; ISBN: 0943088372

Monoclonal Antibodies : A Manual of Techniques by Heddy Zola (September 1987) CRC Press; ISBN: 0849364760

Immunochemistry in Practice by Robin Thorpe (Contributor), Alan P. Johnstone 3nd edition (January 15, 1996) Blackwell Science Inc; ISBN: 0865426333